DOCUMENT RESUME

ED 366 164 EC 302 768

AUTHOR	Ysseldyke, James E., Ed.; Thurlow, Martha L., Ed.
TITLE	Views on Inclusion and Testing Accommodations for Students with Disabilities. Synthesis Report 7.
INSTITUTION	Minnesota Univ., Minneapolis. Coll. of Education.; National Association of State Directors of Special Education, Washington, D.C.; National Center on Educational Outcomes, Minneapolis, MN.; Saint Cloud State Univ., MN.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	Sep 93
CONTRACT	H159C00004
NOTE	69p.; For other reports in this series, see EC 302 769-773.
AVAILABLE FROM	University of Minnesota, National Center on Educational Outcomes, 350 Elliott Hall, 75 E. River Rd., Minneapolis, MN 55455 ($15).
PUB TYPE	Collected Works - General (020)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Adaptive Testing; Deafness; *Disabilities; Educational Assessment; Educational Philosophy; *Educational Policy; Elementary School Students; Elementary Secondary Education; Hearing Impairments; *Mainstreaming; Outcomes of Education; Secondary School Students; *Student Evaluation; Testing Problems; *Testing Programs
IDENTIFIERS	Arizona; Inclusive Schools

ABSTRACT

This monograph provides an overview of the issues surrounding inclusion and testing accommodations for students with disabilities and presents six papers discussing the issues. "Including Students with Disabilities in Systemic Efforts To Measure Outcomes: Why Ask Why?" (Bob Algozzine) argues that excluding any student from testing violates the spirit and practice of full inclusion. "Inclusion and Adaptation in Assessment of Special Needs Students in Arizona" (Paul H. Koehler) describes the Arizona Student Assessment Program, which has tried to include nearly all students with Individualized Education Programs by developing modified (mediated) forms of assessments. "Inclusion of Children and Youth Who Are Hearing Impaired and Deaf in Outcomes Assessment" (Barbara L. Loeding and Jerry B. Crittenden) contends that accommodations and participation in state and national testing should depend upon the deaf student's primary mode of communication, prior use of an interpreter, and functioning level or amount of hearing loss. "Inclusion and Accommodation: 'You Can Tell What Is Important to a Society by the Things It Chooses To Measure'" (Jack Merwin) points out that excluding students with disabilities from state and national testing affects group averages less than excluding other subgroups, such as children from low socioeconomic status groups. "Consequences and Incentives: Implications for Inclusion/Exclusion Decisions Regarding Students with Disabilities in State and National Assessment Programs" (Daniel J. Reschly) explores three inclusion/exclusion policy alternatives. "Inclusion and Accommodation in Assessment at the Margins" (Maynard C. Reynolds) suggests that acceptable test results may be gathered from 95 percent of pupils and that the other 5 percent (students in special education) may be assessed through other means. (Each paper contains references.) (JDD)

ED 366 164

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

EC 302768

BEST COPY AVAILABLE

Synthesis Report 7

Views on Inclusion and Testing Accommodations for Students with Disabilities

Edited by:
James E. Ysseldyke and Martha L. Thurlow

National Center on Educational Outcomes

The College of Education
UNIVERSITY OF MINNESOTA

September, 1993

The National Center on Educational Outcomes (NCEO), established in 1990, works with state departments of education, national policy-making groups, and others to facilitate and enrich the development and use of indicators of educational outcomes for students with disabilities. It is believed that responsible use of such indicators will enable students with disabilities to achieve better results from their educational experiences. The Center represents a collaborative effort of the University of Minnesota, the National Association of State Directors of Special Education, and St. Cloud State University.

The Center is supported through a Cooperative Agreement with the U.S. Department of Education, Office of Special Education Programs (H159C00004). Opinions or points of view do not necessarily represent those of the U.S. Department of Education or Offices within it.

NCEO Core Staff:

Robert H. Bruininks
Kevin S. McGrew
Dorene L. Scott
James G. Shriner
Gail E. Spande
Martha L. Thurlow, assistant director
James E. Ysseldyke, director

Additional copies may be ordered for **$15.00**. Please write:

Publications Office
NCEO
350 Elliott Hall
75 East River Road
University of Minnesota
Minneapolis, MN 55455

Table of Contents

Introduction 1

Including Students with Disabilities In Systemic Efforts to Measure Outcomes: Why Ask Why?
Bob Algozzine 5

Inclusion and Adaptation in Assessment of Special Needs Students in Arizona
Paul H. Koehler 11

Inclusion of Children and Youth who are Hearing Impaired and Deaf in Outcomes Assessment
Barbara L. Loeding and Jerry B. Crittenden 19

Inclusion and Accommodation: "You can tell what is important to a society by the things it chooses to measure"
Jack Merwin 30

Consequences And Incentives: Implications For Inclusion/Exclusion Decisions Regarding Students With Disabilities In State And National Assessment Programs
Daniel J. Reschly 35

Inclusion and Accommodation in Assessment at the Margins
Maynard C. Reynolds 47

Introduction

Two important challenges currently face assessors who are striving for assessment programs that provide complete information on all students in the educational system. These challenges are particularly important for national and state data collection programs because information derived from these assessments contribute to educational policy decisions that potentially have significant impact on students with disabilities. One of the challenges is **inclusion**. Specifically, this challenge involves questions about who should be included in (or excluded from) assessments, how the decision is made, and who makes the decision. The second challenge is **accommodations**. Specifically, this challenge involves questions about what modifications can be made in assessment materials and/or procedures that still allow valid assessment results to be obtained.

These two challenges -- Inclusion and Accommodation -- were the ones described as most challenging for assessors by the Association of State Assessment Personnel (ASAP). In January of 1992, members of the association provided their opinions about important technical issues in assessing outcomes for students with disabilities. Thirteen issues initially were identified:

- definitions
- data quality
- equity
- sampling methodology
- data aggregation
- test standardization
- cross sectional versus longitudinal assessment
- instrument adaptation
- validity
- reliability
- range of items
- out-of-level testing
- feasibility of special studies

Discussions with ASAP members, however, indicated that these topics boiled down to the two key challenges of inclusion in assessment and accommodations in testing. Since the January meeting, the National Center on Educational Outcomes (NCEO) has been involved in several activities to further explore these challenges. This monograph represents the culmination of an effort to obtain expert opinions about these challenges.

Before introducing these opinion papers, we provide a brief overview of the issues surrounding the challenges of inclusion and testing accommodations. These will be followed by an overview of the papers.

Inclusion in Assessments

Currently there is considerable variability in the inclusion of students with disabilities in national and state assessment programs (McGrew, Thurlow, Shriner, & Spiegel, 1992). The most recent estimates of exclusion rates are 40-50% for national programs (McGrew et al., 1992), and from less than 10% to greater than 90% for state assessment programs. In both national and state assessment programs, it appears that most exclusion decisions are made at the local level where criteria may be interpreted in different ways.

Inconsistent use of criteria is probably just one contributor to the extreme variability in exclusion rates. In several national data collection programs criteria are vague. In many states, such criteria do not exist. When guidelines exist, the most frequently mentioned criteria include:

- Time spent in general education settings

- Courses for which the student is mainstreamed
- Match between instructional objectives and assessment objectives

But these are only the few that are most common. Scores of different criteria and combinations of them actually are used by states (see Thurlow, Ysseldyke, & Silverstein, 1993).

Testing Accommodations/Modifications

Like inclusion/exclusion practices for students with disabilities, practices in allowing test accommodations are quite variable. Currently, there is no agreement on the legal implications of recent laws, such as the Americans with Disabilities Act, for testing modifications in many assessment situations. Likewise, written guidelines on accommodations in testing are extremely variable in what they allow. What is recommended in one state may be expressly prohibited in another. Most accommodations that are made involve altering the presentation format or the allowed formats for responding. Changes in settings of assessments and the timing or scheduling of assessments are also common types of accommodations. Yet, there are other types as well, such as out-of-level testing. The controversy over acceptable accommodations extends to minimum competency, certification, and licensure tests also.

Underlying this uncertainty about acceptable accommodations is a lack of data on testing modifications. Limited research has been conducted on the effects of various modifications. Furthermore, this lack of knowledge makes it difficult for test users and developers to follow the existing standards for educational and psychological tests, which mandate that any changes in testing procedures require additional validation of the test. Consequently, concerns about the validity and reliability of modified procedures are prevalent, especially among the publishers of tests.

Overview

Ultimately, we need to develop a set of guidelines that address inclusion/exclusion and accommodation issues. As a first step in developing guidelines, we asked experts in the field to write papers on technical issues in the assessment of educational outcomes for children and youth with disabilities. Some of the questions that these contributors were asked to address were:

- What kinds of assessment now go on in the states?
- What are the benefits and drawbacks of each type?
- What are the ways in which decisions might be made about inclusion of students with disabilities in state and national assessments?
- What are the advantages and disadvantages associated with each approach?
- What are the ways in which decisions might be made about the adaptation of state and national measures for students with disabilities and who should make such decisions?
- What criteria/guidelines might be reasonable to use in making decisions about whom to include in state and national assessments and the kinds of adaptations that ought to be permitted?

The following individuals contributed papers on these issues:

Bob Algozzine
Professor of Teaching Specialties
University of North Carolina-Charlotte

Paul Koehler
Arizona Department of Education

Barbara Loeding and Jerry Crittenden
University of South Florida

Jack Merwin
Professor Emeritus of Educational Psychology
University of Minnesota

Daniel Reschly
Distinguished Professor of Psychology
Iowa State University

Maynard Reynolds
Professor Emeritus of Educational Psychology
University of Minnesota

The papers reflect the diverse arrays of opinions about the inclusion and accommodation issues. The order in which they are presented is alphabetical and has no other significance.

Algozzine argues that excluding any student from testing violates the spirit and practice of full inclusion. For the same reason, he is against the practice of substituting progress on an Individualized Educational Plan (IEP) for state and national testing results. Finally, he thinks that any accommodations or modifications offered to one student should be available to all students.

Koehler views exclusion as being problematic in that removing students in special education from the "accountability track" (state and national testing) also results in removal from the "curriculum track" since IEP goals are often unrelated to regular education goals. He describes the approach used in the Arizona Student Assessment Program, which has tried to include nearly all students with IEPs by developing modified (mediated) forms of the assessments. Arizona also developed a set of guidelines on "mediated assessment," with accommodations intended to mirror, as closely as possible, "normal" instructional adaptations.

Loeding and **Crittenden** focus on the technical issues of assessing educational outcomes of students who have hearing impairments or deafness. They say that accommodations and participation in state and national testing should depend upon the deaf student's (a) primary mode of communication, (b) prior use of an interpreter, and (c) functioning level or amount of hearing loss. They also suggest that full inclusion or full exclusion of students with hearing impairments violates PL 94-142, which requires nondiscriminatory testing.

Merwin writes that it is acceptable to exclude children with disabilities from state and national testing because students in special education comprise such a small number of students that their exclusion will not affect state and national comparisons. Moreover, excluding students with disabilities affects group averages less than excluding other subgroups, such as children from low socioeconomic status groups.

Reschly explains that inclusion decisions historically have varied with specific outcome domains and the stakes of results. Generally, high consequences lead to the unwarranted exclusion of children with disabilities. Some methods of exclusion are subtle and difficult to document, such as encouraging a child to stay home on a testing day, or indicating that an answer sheet was not completed properly and results thus are invalid. Reschly explores the advantages and disadvantages of three inclusion/exclusion policy alternatives for high stakes state and national assessment programs: (1) full exclusion, (2) full inclusion, and (3) allowing the exclusion of 2% of the students. He argues that the implementation of liberal accommodations policies would probably increase the perception of fairness and the assessment programs' credibility. Finally, he suggests methods to improve the overall integrity of assessment programs.

Reynolds favors universal (inclusive) assessment practices for carefully specified, culturally imperative domains: language, mathematics, social skills, and self-dependence. He suggests that acceptable test results may be gathered from 95% of pupils, and that the other 5% (students in special education) may be assessed through alternative testing, adaptive measurement procedures, or teacher judgments.

References

McGrew, K.S., Thurlow, M.L., Shriner, J.G., & Spiegel, A.N. (1992). Inclusion of students with disabilities in national and state data collection programs. Minneapolis, MN: National Center on Educational Outcomes.

Thurlow, M.L., Ysseldyke, J.E., & Silverstein, B. (1993). Testing accommodation for students with disabilities: A review of the literature (Synthesis Report 4). Minneapolis: National Center on Educational Outcomes.

Including Students with Disabilities In Systemic Efforts to Measure Outcomes: Why Ask Why?

Bob Algozzine
University of North Carolina, Charlotte

> *A better education than special class placement is needed for socioculturally deprived children with mild learning problems who have been labeled educable mentally retarded.* (Dunn, 1968, p. 5)

Current organizational practices promote the separation of general and special education students and programs. An elaborate system of assessment and classification has evolved to support the need and conduct of separate systems of education. For many students, this system provides educational benefits that are otherwise unattainable in traditional school programs and special education has come a long way. Parents and other professionals have accomplished significant and important gains in their efforts to bring education for students with disabilities out of the basements, closets, and special classes that too often reflected the diminished perspective assigned to other than mainstream educational experiences. Parents, teachers, policy-makers, and other dedicated professionals can "boast that special education serves students who were formerly excluded from schools and special education has developed into a sophisticated system serving approximately four and a half million students nationwide each school year" (Roach, 1991, p. 1).

Despite all the progress, important concerns continue to be identified and the cry for reform issued more than twenty years ago by Lloyd Dunn echoes today:

> Of great concern is the criticism by some that a disproportionate number of minority students have been placed in special education. Educators have been criticized for unnecessarily labeling and segregating students; for educating students in overly segregated settings; for not providing substantially different methods of instruction; and for limiting curricular options based on students' handicap label. (Roach, 1991, p.2)
>
> And yet, today, as a decade [and, two decades] ago, special education has a host of unsolved problems: Numbers of students with disabilities are increasing, resulting in teacher-pupil ratios that sometimes rival those in general education; children of African-American and Hispanic heritage continue to be over-represented in many special education programs; numerous school districts still do not value reintegration as an important goal, with the result that special education is a terminal assignment in the educational careers of too many students; and evidence is still insufficient that what many special educators do is indeed special and effective. These and other concerns, together with spiraling special education costs in a recession that is squeezing education budgets in state capitols, increases the likelihood that the field and its problems will not be overlooked. (Fuchs & Fuchs, 1992, p. 413)

More than 20 years after Dunn's swan song begging for reform and restructuring of special education, significant problems remain and a new wave of initiatives to improve the system has begun. With the latest efforts to reform special education, "inclusionary schools" and "full-inclusion" have become the catch phrases of this decade.

The goals of full-inclusion are difficult to defile. As described in the New Mexico State Department of Education's Administrative Policy on Full Inclusion (Morgan, 1991), full inclusion means that all children are educated in supported, heterogeneous, age-appropriate, dynamic, natural, child-focused classroom, school, and community environments. Full inclusion means open doors, accessibility, proximity, friends, support, and valuing diversity. Full inclusion means attending a school of choice, attending classes with neighbors and natural peers, and participating in school and community activities that maximize social development of everyone. Schools that practice full inclusion take responsibility for the learning of all students. Full inclusion is given weight by the Individuals with Disabilities Education Act which calls for serving students with disabilities in "least restrictive environments," by Section 504 of the Rehabilitation Act which guarantees people with disabilities access to services provided by federally funded agencies, and by the Americans with Disabilities Act which requires that employers make work facilities readily accessible to and usable by people with disabilities (Ysseldyke, Algozzine, & Thurlow, 1992).

Though not a new idea, the practice of full inclusion has recently impacted personnel concerned with assessment and educational outcomes. Findings that significant numbers of students with disabilities are not included in state reports of pupil performance and national data bases have merely added to the urgency of these concerns (McGrew, Thurlow, Shriner, & Spiegel, 1992). In attempting to justify exclusionary practices, state and national assessment personnel argue that several kinds of decisions must be made when including students with disabilities in state and national assessment reports. First, decisions must be made about when and for what purposes state and national assessments must be given. Second, decisions must be made about the conditions under which students with disabilities should be a part of these state and national assessment efforts. These concerns represent the areas addressed in this paper.

When and What For?

Efforts to assess educational systems flourish in eras of reform and innovation. Probably no single factor has pushed current accountability efforts more than activities that surround the nation's current educational reform strategy and the National Education Goal Panel (NEGP) that has assumed responsibility for monitoring it. Current educational reform initiatives have produced a flurry of state and federal activity focused on identifying indicators of progress toward national educational goals. Developing indicator systems has become big business in the United States (Odden, 1990), with nearly all national and state education agencies becoming more involved in decision making, monitoring, accountability, and measuring educational progress than ever before in the nation's history (McGrew et al., 1992).

The United States has a developing and rich tradition of assessing students' progress as a measure of the overall quality of its educational system (McGrew et al., 1992). Scores on cumulative tests (generally standardized) administered at selected school transitions (e.g., graduation, promotion to third grade) serve as data for decision making and documentation of need for improvements and programs. National data collection programs such as the National Assessment of Educational Progress (NAEP--The "Nation's Report Card"), the National Education Longitudinal Study (NELS) are a few examples of recent and continuing efforts to provide periodic data on the educational status of America's school children. Graduation testing, minimal competency testing, minimal skills diagnostic testing, and end-of-grade or end-of-course achievement testing are examples of when state departments of education administer tests and their reasons for administering them. Excluding students from any of these causes serious concerns when compiling, reporting, and interpreting scores.

Under conditions of national importance, policy decisions should be made on the basis of information reflecting all students. Problems arise when states and federal agencies do different things. Riding on the inclusion bandwagon, this means deciding assessment issues is relatively simple and straight-forward. In full inclusion, any experience available to one student is available to all students. No students are excluded from any activities that are available to neighbors and peers. From an assessment standpoint: Any tests administered to one student as part of state and national assessment practices should be administered to all students. No students are excluded from any activities available to neighbors and peers. Any tests used to make decisions about one student are used to make decisions about all students.

This seemingly simple solution (and, any more complex) creates an apparent assessment quagmire: Professionals with a more traditional, conservative, or technical view believe that if all students are included in state and national assessment efforts, then the picture presented is biased by the performance of students with disabilities. Those with radical, liberal, or practical views argue that if students with disabilities are not included, then a biased picture of local, state, or national performance is presented. And, the dilemma is further complicated when asking "why" students are tested in the first place (e.g., high stakes vs. low stakes).

As is often the case in education, the simple solution is set aside and the practice wants to be complex. Problems arise when states, in the absence of federal or professional direction to do otherwise, do different things. Clearly this is the state of current practice and this apparently simple solution becomes complex when the question of inclusion becomes a technical rather than practical concern.

Who Should be Tested?

Foremost among the technical questions that professionals want to complicate assessment decision making is who should be included in national, state, and local school performance data bases. Among the alternatives are excluding all students from a particular group, excluding some students from representative groups but not others, or including all students. Again, to simplify this area of concern with a full inclusion viewpoint, all means all. Excluding any students violates the spirit and force of inclusionary practices and, clearly, when any students are excluded, the effects vary greatly depending on the sampling process. Perhaps no better illustration of this exists than one presented by continuing abuse of Scholastic Aptitude Test (SAT) performance across states as a measure of educational quality within states.

Indictments of using state-to-state variation in standardized test performance as indicators of quality and equality in education are not new. Womer (1983) noted the hazards for those in the testing community: "the SAT is *not* an agency of educational accountability; and the SAT is *not* designed to sample curricular attainments of high school seniors nationwide; and the SAT is *not* administered to random samples of high school seniors" (p. 4, emphasis in original). In 1984, Powell and Steelman found that "most state variation [in overall SAT performance scores] is a function of one factor, the percentage of eligible students taking the exam" (p. 408). To "account for some of the variation in state means for the SATs" Page and Feifs (1985) used percentage of minority students, pupil-teacher ratio, teacher salaries, state employment rates, and several other predictor variables in their regression analyses. They found that "the most powerful influence on the state means, of course, [was] the percentage of high school graduates who [took] the SAT exams" (p. 310). Most recently, Wainer (1989) discussed problems related to drawing conclusions from samples in which selection criteria are not known (e.g., students who choose to take the SATs). Using a re-analysis of data provided by earlier investigators (i.e., Page & Feifs, 1985; Powell & Steelman, 1985), he found a moderate correlation (0.5) between state rankings, but considerable variance relative to standings of some individual states (e.g., New Mexico was first and 30th in separate analyses). A distinguished set of colleagues commented and generally supported Wainer's position (cf. Allen & Holland, 1989; Birnbaum & Mellers, 1989; Heckman, 1989; Rosenbaum, 1989; Rubin, 1989; Speed, 1989; Wachter, 1989) that variances in sample characteristics create significant problems when comparing performances across sampling groups.

The complexities created by excluding students from local, state, and national testing programs bear this weight as well. If all students are included, differences in performance across comparison groups are due to naturally-occurring differences in characteristics of comparison groups. If some students are excluded, differences in performance across comparison groups are *not* due to naturally-occurring differences in characteristics of comparison groups. For example, if one state excludes all and another excludes some students with learning disabilities, reporting and comparing outcomes across the states becomes meaningless. Considering the combinations of student types that may be included or excluded in such practices illustrates the complexity of problems created by selective inclusionary practices in assessment of outcomes.

Under What Conditions?

Additional technical questions that many believe must be addressed concern how to modify assessment practices to include students with disabilities. The strictest interpretation permits no modifications and requires any students included in local, state, or national assessments to take standard versions of tests being used in decision making. Options under modified assessment practices include permitting alternative assessment procedures to accommodate the needs of some students and incorporating test modification into local, state, and national assessment practices. From the full inclusion perspective, if standard versions of tests are administered to some students, they should be administered to all students and any accommodations or modifications offered to one student should be offered to all students.

Wildemuth (1983) described several approaches taken to accommodate students with disabilities in state assessment programs: Excluding them from the testing requirement, using an individualized education program (IEP) as basis for decision making, and establishing different standards. Each fails the full-inclusion test. First, excluding students from the testing requirement because of a disability violates the spirit and purpose of inclusion. Permitting the IEPs to substitute for performance on tests taken by general education students and establishing different standards for performance are discriminatory, selective practices that also violate the sentiments of full inclusion.

Large-print versions of tests, video-cassette versions of tests, Braille editions, audio-cassette versions of tests, extended time periods, marking in test booklets, using proctors to record answers, using typewriters or word processing computer programs, selecting test items that best evaluate educational objectives, developing alternative tests, and using alternative testing procedures (e.g., reading the test to a student with reading problems) are among test modifications permitted in some states (North Carolina Department of Public Instruction, 1992). Many of these changes improve the performance of students with disabilities and most would improve the performance of any student (Beattie, Grise, & Algozzine, 1983). In North Carolina, an IEP committee decides on the modifications that will be allowed for particular students (Amos, 1980; North Carolina Department of Public Instruction, 1992), but students participating in modified assessment procedures are "exempted from the Annual Testing Program for purposes of accountability" (p. 2).

Clearly these are discriminatory practices from a full-inclusion point-of-view. Any modification allowed for one student should be allowed for all students. Excluding students for any reason violates the spirit and intent of inclusion in the first place and causes serious concerns when reporting and interpreting scores. The availability and use of test modifications and alternative procedures brings the assessment quagmire back to its origin: When and for what purposes are test modifications and alternative assessment procedures to be used?

A Step In Some Direction

Solving the problems apparent when considering inclusion of students with disabilities in state and federal assessments of education outcomes will not be easy. Treating them as part of a broader, more inclusive view of education offers an alternative that at least reduces the likelihood of "marching in place." And, while it is tempting to approach questions related to full inclusion from a technical basis, they are not technical questions. While it is tempting to argue that it shouldn't be done until benefits of doing it have been proven, it is not a problem that requires cost-benefit solutions. While it is tempting to argue that test modifications should not be permitted because they violate the technical boundaries of psychometric practice, again these should not be treated as technical considerations. It is better to view these problems, in the purest sense of what is going on today, from the context that all tests and testing procedures lack perfect technical adequacy. In an imprecise domain, laboring under the pursuit of unattainable ideals is like rolling boulders up a mountain or continually marching in place. A simple solution in cases such as these is often to simply take a step in some direction. Toward this end, consider the following:

To improve assessment of outcomes in America' s schools, professionals should avoid any practices that produce, encourage, foster, or facilitate separation among student groups. All students should be expected to take all tests and any modifications permitted for any test or any assessment procedures should be permitted for all tests, all assessment procedures, and all students. If society expects and allows accommodations in ordinary life, testing agencies should expect and allow them as well. Scores obtained as part of state and national efforts to assess performance should be reported in fully aggregated (including all students) and disaggregated (by appropriate student group) forms.

To do less creates more problems than solutions and simply doesn't make sense as sound educational practice.

References

Allen, N. L., & Holland, P. W. (1989). Exposing our ignorance: The only "solution" to selection bias. Journal of Educational Statistics, 14, 141-145.

Amos, K. M. (1980). Competency testings: Will the LD student be included? Exceptional Children, 47, 194-197.

Beattie, S., Grise, P., & Algozzine, B. (1983). Effects of test modifications on the minimum competency test performance of learning disabled students. Learning Disability Quarterly, 6, 71-77.

Birnbaum, M. H., & Mellers, B. A. (1989). Mediated models for the analysis of confounded variables and self-selected samples. Journal of Educational Statistics, 14, 146-158.

Dunn, L. M. (1968). Special education for the mildly retarded -- Is much of it justifiable? Exceptional Children, 35, 5-22.

Fuchs L. S. & Fuchs, (1992). Editorial: special education's wake-up call. Journal of Special Education, 25, 413-414.

Heckman, J. J. (1989). Causal inference and non random samples. Journal of Educational Statistics, 14, 159-168.

Morgan, A. D. (1991). Memorandum: New Mexico State of Department of Education's Administrative Policy on Full Inclusion. Special Net (Posted January 8, 1992).

McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. (1992). Inclusion of students with disabilities in national and state data collection systems. Minneapolis, MN: National Center on Educational Outcomes.

North Carolina Department of Public Instruction. (1992). Guidelines for testing exceptional students. Raleigh, NC: Author, Research and Development Services.

Odden, A. (1990). Educational indicators in the United States: The need for analysis. Educational Researcher, 19 (5), 24-28.

Page, E. B., & Feifs, H. (1985). SAT scores and American states: Seeking for useful meaning. Journal of Educational Measurement, 22, 305-312.

Powell, B., & Steelman, L.C. (1984). Variations in state SAT performance: Meaningful or misleading? Harvard Educational Review, 54 (4), 389-412.

Roach, V. (1991). Special education: New questions in an era of reform. Issues in Brief, 11 (6), 1-7.

Rosenbaum, P. R. (1989). Safety in caution. Journal of Educational Statistics, 14, 169-173.

Rubin, D. B. (1989). Bugs, lacunae, and the Minnesota/DC effect: A discussion of Wainer's "Eelworms, bullet holes, and Geraldine Ferraro." Journal of Educational Statistics, 14, 175-178.

Speed, T. P. (1989). Contribution to discussion of "Eelworms, bullet holes, and Geraldine Ferraro." Journal of Educational Statistics, 14 179-181.

Wachter, K. W. (1989) Statistical adjustment: Comments on H. Wainer's "Eelworms, bullet holes, and Geraldine Ferraro." Journal of Educational Statistics, 14 183-186.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. Journal of Educational Statistics, 14 121-140.

Wildemuth, B.N. (1983). Minimum competency testing and the handicapped. Princeton, NJ: ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Womer, F. B. (1983) Congratulations, Mississippi! Shame on you, Rhode Island! (Editorial). Educational Measurement, 2, 4.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (1992). Critical issues in special education. Boston, MA: Houghton Mifflin.

Inclusion and Adaptation in Assessment of Special Needs Students in Arizona

Paul H. Koehler
Arizona Department of Education

The inclusion of special needs students in statewide assessment programs is a topic of considerable discussion in states such as Arizona which are undergoing revisions of their statewide programs. It has been customary practice for special needs students, especially those who have Individualized Education Programs (IEPs), to be systematically excluded from statewide assessment programs due in part to the difficulty in developing appropriate assessments for such students. Arizona has taken steps to change this practice and include nearly all students with IEPs in the new Arizona Student Assessment Program (ASAP).

Standardized testing has long been the anchor of statewide assessment programs. The reasons for this are well-known. The "back to basics" movement encouraged the use of tests that sought to measure basic skills in an efficient manner for all students. Arizona, for example, has had a statutory requirement since 1980 that, in effect, mandated the use of a norm-referenced standardized test for all regular education students grades 1-12 each year. This extreme example of testing of the "basics" for reasons of accountability was repeated to a somewhat lesser degree in most states during the 1970s and 1980s. While the results of such testing programs produced mountains of data at the school district and state level, as well as questionable test practices and interpretations of the data (Haladyna, Haas, & Nolen, 1989), the standardized test programs also excluded special education students. The exclusion was based on the belief that the features of standardized testing, including strict adherence to time of administration, prohibition of student-teacher interaction during test administration and computerized scoring, was not concurrent with the instruction and assessment process used in the instruction of most special education students and would therefore be a detriment rather than a benefit for them. The literature on this subject consistently yields the opinion that traditional assessment often does nothing more than confirm the poor skill level of special education students (Guerin, 1991). For this reason language has often been added to statewide assessment mandates removing students with IEPs from all of the testing.

The exclusion of special education students from the statewide assessment programs solved one problem but created another. The negative impact often experienced by students being tested with norm-referenced tests was removed but the information available about the academic performance of special education students was lost as well. In the example of Arizona, and probably in other states, the removal of special education students from the "accountability track" also resulted, to a large degree, in their removal from the "curriculum track," those learning expectations that guided the instruction of regular education students. As district and schools in Arizona have reviewed the quality of students' IEPs, it is often the case that the instructional goals are unrelated to the more ambitious goals evident for regular education students. This results in the special education student becoming more and more isolated from the mainstream instructional program rather than having an alternate course being charted for reaching competence in the mainstream subject area content. While the blame for such practice cannot be attributed solely to statewide standardized testing programs and special education students' exclusion from them, there is a strong case to be made that the absence of anchoring assessments to statewide standards has contributed significantly to the problem.

In the past few years many states have made considerable progress toward setting high curriculum standards and developing assessment systems that are more closely aligned with good instructional process. Raizen and Kaser (1989) believe that assessment performs an important role in curriculum change. They note that "assessment can be a powerful tool for reform, since changing the nature of assessment can lead to changing the nature of instruction" (p. 720).

Arizona embarked on this course in 1988 with the development of the Arizona Student Assessment Program (ASAP), which became law a year later. ASAP redefined what students should know in all of the content areas at three key benchmark grades (3, 8, 12) and established a performance assessment system for the skills in the content areas (reading, mathematics and writing have been completed, science and social studies are under development). The new curriculum standards expressed, in performance terms, the learning goals for all students including special education students. The goals provided for the first time, a set of high quality expectations for all teachers, schools and districts that could be used as a base from which to develop instructional activities for all students. The expectations were addressed to all students, allowing special needs students (including both special education and gifted) to have their programs modified from a realistic base of expectations. Previously all that was available to the schools were the skills targeted to the standardized test which, although not intended to serve as a state curriculum, became the student, school and district focus of instruction across states such as Arizona. Both instruction and assessment developed at the district and school level since 1980 began to look like the standardized tests that served as their influence. The new standards and assessments in ASAP have already started to influence local school and district thinking about curriculum and assessment decisions in a way never before seen with only the standardized test as the standard for achievement.

From its beginning, ASAP was designed to be inclusive of special needs students. The broad-based input on ASAP received from across Arizona included that of parents and teachers of special needs students, including limited English proficient students, special education students and students with disabilities. ASAP represented the first intentional involvement of all of these students in a program that would attempt to emphasize what all students should do rather than the all-too-familiar approach with special needs students of what they cannot do.

ASAP included assessment geared toward performance of the curriculum standards at the targeted grades of 3, 8, and 12. But teachers also requested some consideration of assessment development at grades prior to the targeted grades. For this reason two forms of the assessments (one, a grade level or two easier than the targeted grade level assessment, and the other, two or three grade levels easier) were developed. The importance of these alternate forms is that they were developed as measures of the same high level curriculum standards measured by the targeted grade assessment. This allowed, for the first time, students in special education programs or those with disabilities who are working on a learning pace slower than their regular education counterparts, to be assessed on the same skills but on a less rigorous basis. The alternate forms of the assessment, in effect, provided a statewide program of "materials and methods modification" often employed by teachers of special needs students. The development and dissemination of these assessments sent a strong message throughout the Arizona schools that ASAP was for everyone.

The next step in the effort to include special needs students in ASAP occurred during the Spring, 1992 statewide piloting of the assessments at grades 3, 8 and 12. The decision was made that all students except those most severely impacted students with IEPs that specifically prohibited any form of testing, would be included in the piloting of assessments. Since the pilot administration of the assessment was truly a "low stakes" experience with no public reporting of the results below aggregated state level data, it allowed educators with experience in the instruction and testing of special needs students to provide guidance on how the administration should be conducted. A set of guidelines for "mediated assessment" was developed which, in reality, mirrored the kinds of instructional adaptations usually seen in the instructional methodology employed by teachers of special needs students or students with disabilities. The purpose of the pilot testing allowing mediation was to gather data on its effect on student performance and especially on the validity and reliability of assessments administered with mediation.

The Department of Education issued the guidelines shown in Figure 1 for mediation after consulting with knowledgeable educators about what is possible. For purposes of the pilot study, it was important to know which students were administered mediated assessments and the type of

Figure 1

Special Education Students or Students with Disabilities Definitions and Guidelines for Mediation

Definition

Students who are special education students included under the Individuals with Disabilities Education Act (IDEA) whose individual Education Program (IEP) states that mediation is required

or

Students who are not considered special education students under IDEA but are considered to be students with disabilities and are covered under section 504 of the Vocational Education Rehabilitation Act.

Guidelines for Mediation

The following may be used:

1. Providing flexible scheduling
    - Extending the time allotted to complete the assessment.
    - Administering the assessment in several sessions.
2. Providing a flexible setting
    - Administering the assessment individually in a separate location
    - Administering the assessment to a small group in a separate location
    - Providing special lighting
    - Providing adaptive or special furniture
    - Providing special acoustics
    - Administering the assessment in a location with minimum distractions
3. Revising assessment directions
    - Reading directions to student
    - Simplifying language in directions
    - Highlighting verbs in instructions by underlining
    - Providing additional examples
4. Providing assistance during the assessment
    - Reading questions and content to student
    - Signing questions and content to students
    - Taking dictation
5. Using aids
    - Visual magnification devices
    - Auditory amplification devices
    - Auditory tape questions
    - Masks or markers to maintain place
    - Tape recorder
    - Typewriter or word processor
    - Communication device
    - Calculator
    - Abacus
    - Arithmetic Tables

mediation that was employed. This information was gathered on the scannable form for scorers as illustrated in Figure 2.

It is worth noting that the nature of the assessments in ASAP require that the scoring be done by "real people" rather than by computers scanning students' bubble sheets. Typically, ASAP requires students to use a variety of modes of response including writing, illustrating and graphing. Scorers judge the quality of student responses based on rubrics and examples of correct answers from the students (anchor papers). Student scores are then transferred onto the scan sheets illustrated in Figure 2.

For the pilot study it was essential to know whether there was a mediated administration of the assessment, the reason for the mediation and the type of mediation provided (see Figure 2). This information then became the subject of a special study to determine the number and type of mediation used with students in grades 3, 8 and 12. The effect of the mediation is also under study. The decision was made that the mediated assessments were to be scored along with the regular assessments with no special consideration or even awareness of their origin on the part of the scorers. However, after this first scoring, the mediated assessments of the special needs students were rescored by individuals with special training and sensitivity to the reasons for mediation for the examinee. A study was conducted to determine whether the training of the scorer had an impact on the judgments made during the scoring process.

All ASAP assessments with the special education classification encountered by Arizona Department of Education personnel at the scoring sites in Spring 1992 were double-scored by special education teachers. The experimental condition imposed was scorer knowledge of basic student demographic information and special classifications during the second scoring. The population was a convenience sample, systematically excluding cases where an appropriate special education instructor was not available at the site to perform the second scoring.

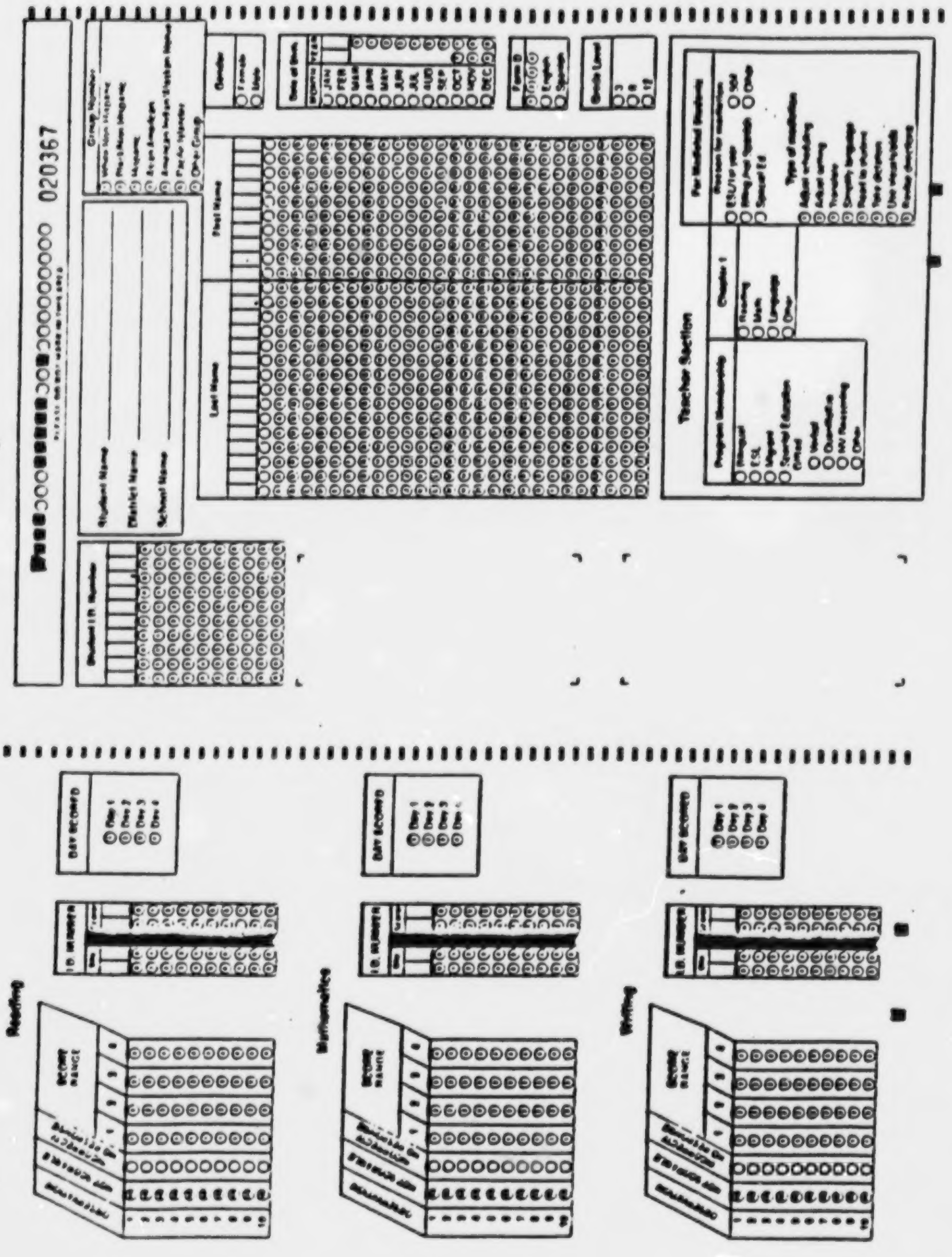
The estimate of the special education population participating in ASAP in 1992 is that over 3,300 student booklets were scored blindly. Of that number, 183 student booklets were captured and double-scored at the 15 ASAP scoring sites, with 110 of the double-scored booklets matched based on student name. Information on the special study is presented in Table 1.

The results of the study showed that, overall, the correlation between the blind scoring and second scoring of the assessments was 0.965 (110 valid cases). Seventy-two cases of the paired scores were not more than one point apart; 45 cases had exact matches. Though the sample was not random, the demographic information roughly matched the state special education population that participated in ASAP. The mean difference between the scores (first score minus second score) was -0.082, indicating that virtually no clear direction of bias was present in the study. (See Table 1 and Figure 3.)

The Arizona Student Assessment Program is the first attempt in Arizona to conduct a large-scale assessment program that is inclusive of all students. The decision by the legislature, state board and Department of Education to allow the kinds of mediation strategies required to make assessments accessible to special needs students was made with full knowledge of the traditional concerns such strategies have had with respect to the validity and reliability of the assessments. That risk is acceptable at this point knowing that the studies described above will contribute to the understanding of the real importance of the issues of mediation. Those working on the development of ASAP agree that knowing why a student could not be judged competent on a skill is an important issue. Zigmond, Vallecorsa, and Silverman (1983) suggest two alternative hypotheses for student errors: (1) Errors made because, although the skill was known, the test conditions prevented a demonstration of the skill; or (2) Errors made because the skill was not known. It is the difference

Figure 2

Scannable Form Teacher Section With Information on Mediation



BEST COPY AVAILABLE

Table 1

ASAP Special Education Special Study Results

Special Education Population Participating in ASAP

	Matched Group		State Population	
	N	%	N	%
White	71	70.3	1979	63.0
Black	4	4.0	143	4.5
Hispanic	21	20.8	729	23.2
Asian	1	1.0	8	0.3
Amer. Indian/Alaskan Native	3	3.0	236	7.5
Pacific Islander	na	na	9	0.3
Other	1	1.0	39	1.2
Total	101	29.8	3143	100.0
Male	67	65.7	2185	67.0
Female	35	34.3	1074	33.0
Total	102	100.0	3259	100.0

Absolute Value Differences between the First and Second Scoring

	N	%
exact match	45	40.9
1 point difference	27	24.5
2 point difference	25	22.7
3 point difference	9	8.2
4 point difference	3	2.7
5 point difference	0	0.0
6 point difference	0	0.0
7 point difference	1	0.9
Total	110	99.9

Summary Statistics of the Difference between Matched Cases

(First Score - Second Score)

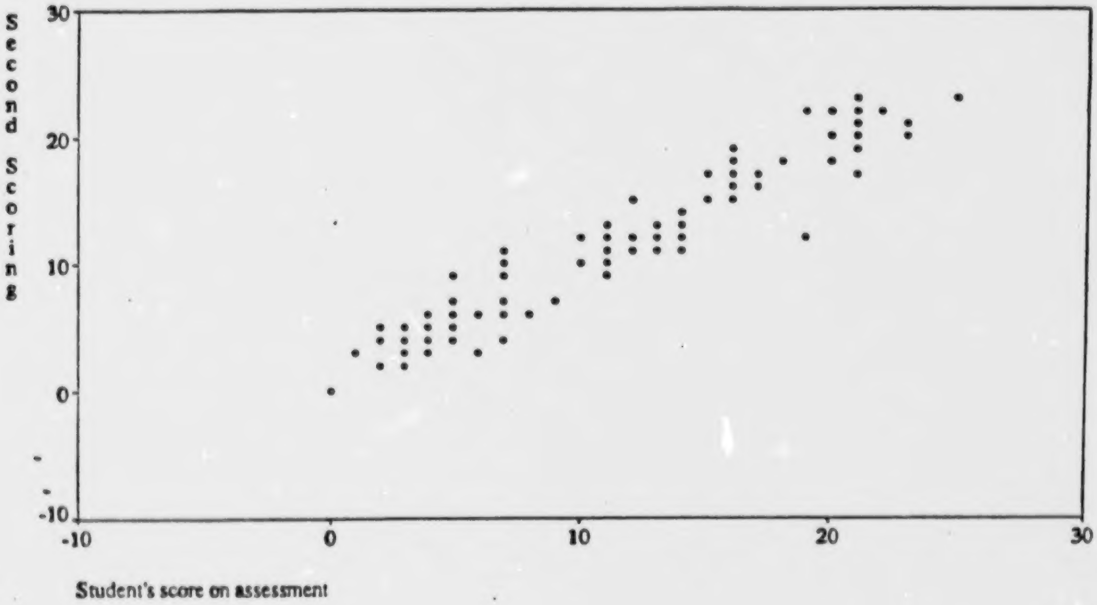
Mean	Std.Dev.	Min.	Max.	Valid N
- 0.082	1.671	- 4	7	110

Pearson Product-Moment Correlation between Blind Score and Second Score

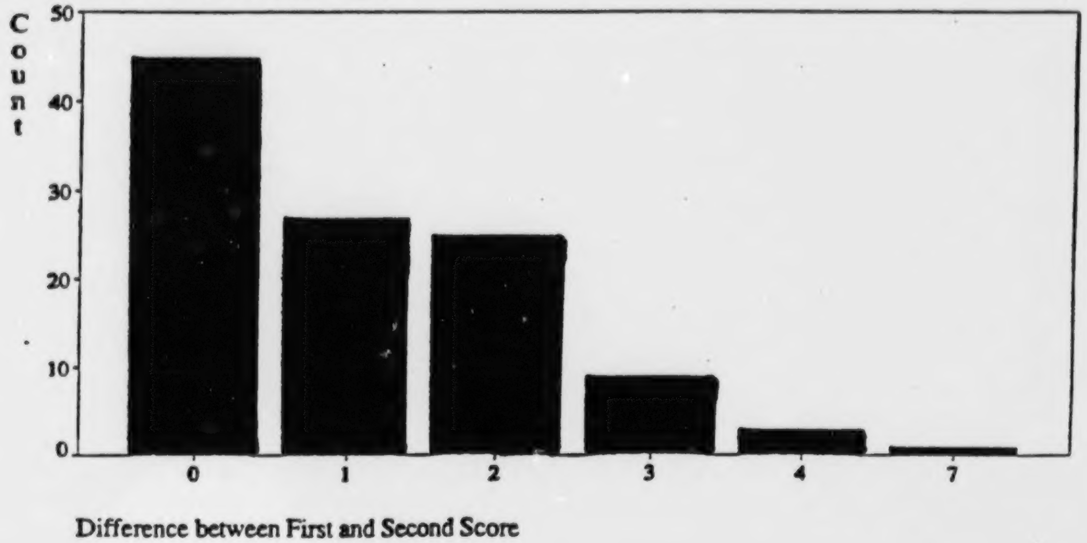
r	signif.
0.965	0.000

Figure 3

Relationship Between Blind Scoring And Scoring with Knowledge About Student



Distribution of the Absolute Value of the Difference between Scores



between the first and second reasons that may be addressed by the administration modifications in ASAP. It is too soon to predict the influence of ASAP on the educational programs and success of special needs students. The goal of making the statewide curriculum goals and related assessments accessible to this population of students is clearly tied to the belief by the educators who work with these students that this is exactly what is needed to improve success for them in school. What can be said with some certainty at this point is that by the very act of including all special needs students in ASAP and providing means for them to access the curriculum goals and assessments, a strong message has been delivered to the schools in Arizona that for once educational practice is consistent with the educational rhetoric regarding the inclusion of all students within the community of learners.

References

Haladyna, T., Haas, N. & Nolen, S. B. (1989). Technical Report 89-1: Test score pollution. Phoenix: Arizona State University, West Campus.

Guerin, R. (1991). Critical step in curriculum reform: Regular education materials and special needs students. Sacramento: California State Department of Education.

Raizen, S. A. & Kaser, J. S. (1989). Assessing science learning in elementary school: Why, what and how. Phi Delta Kappan, 70, 718-722.

Zigmond, N., Vallecorsa, A., & Silverman, R. (1983). Assessments for instructional planning in special education. Englewood Cliffs, NJ: Prentice-Hall.

Inclusion of Children and Youth who are Hearing Impaired and Deaf in Outcomes Assessment

Barbara L. Loeding and Jerry B. Crittenden
University of South Florida

America 2000 (the Bush reform initiative) and now Goals 2000 (Clinton's reform inititative) call for national and state assessment of all students. In order to do this, we must address the technical concerns and the extent to which students with disabilities will be included in the proposed national assessment. Specifically, this paper addresses the technical issues of assessing educational outcomes of children and youth who are hearing impaired and who are Deaf. In the field of deafness, there is continuing debate over correct terminology. This article will use a broad definition of the term "deaf" to refer to all persons with hearing impairments, including those who are hard-of-hearing, those deafened later in life, those who are profoundly deaf, etc. This definition is the one which has been adopted by the National Commission on Education of the Deaf (Commission on Education of the Deaf, 1988). In certain instances, the word, "Deaf" has been capitalized to indicate reference to a separate culture of individuals whose primary language is American Sign Language (ASL) and who identify themselves with Deaf Culture (just as "Spanish" would be capitalized).

When considering how to make decisions about the inclusion of deaf students in standardized state and/or national assessments, there are several areas to consider. The first area is the necessity of language-fair assessments, as recommended by the National Commission on the Education of the Deaf in 1988 and the implications for students with hearing loss. It is important to define and clarify the variety of communication modes used by different deaf students. The second area is establishment of conditions under which students who are deaf should be included or excluded from proposed state and national assessment efforts. The process should include establishment of guidelines or criteria for determining whether inclusion in standardized testing or inclusion in an adapted assessment is appropriate. The third area is determinination of who should make the decisions about whether a child meets the guidelines for inclusion in the standardized assessment or for taking the adapted or modified assessment. And the final area is determination of what adaptations should be provided for students not meeting the guidelines for inclusion in standardized assessments.

Language-Fair Access for Deaf Students

The difficulty of evaluating bilingual students in a fair manner is widely recognized. When a bilingual student is suspected of having a learning disability or a speech and language impairment, evaluation is even more complex because of the difficulty of distinguishing behaviors associated with acquiring a second language from those associated with the suspected disability (ERIC, 1989). Since Deaf students, for the most part, are acquiring English as a second language (Educational Testing Service, 1991; Jordan & Karchmer, 1986), this paper makes the assumption that evaluation of hearing impaired and/or Deaf students is complicated for the same reason. For these students, their primary communication mode is either a visual-spatial, natural sign language used by members of the American Deaf community called American Sign Language (ASL) or a manually coded form of English (Signed English, Pidgin Sign English (PSE), Seeing Essential English (SEE 1), Signing Exact English (SEE 2), or Sign-Supported Speech/English.

Therefore, traditional paper-and-pencil tests are inaccessible, invalid and inappropriate to the deaf student because the tests are in written English only. To complicate the situation further, the ethnicity of thirty-seven percent of deaf students is other than Anglo-American (Christensen & Delgado, 1993). The population of deaf students whose family uses a language other than English or ASL as the primary mode of communication is growing every year (Christensen & Delgado, 1993).

In 1988, the National Commission on Education of the Deaf (COED) was formed to study the quality of education programs furnished to deaf individuals. The membership of this Commission, which had a deaf majority, presented 52 recommendations in their final report. Among the recommendations were following:

> The Department of Education should monitor states to ensure that the evaluation and assessment of children who are deaf be conducted by professionals knowledgeable about their unique needs and able to communicate effectively in the child's primary mode of communication. (COED Recommendation #12)
>
> The Department of Education should take positive action to encourage practices under the Bilingual Education Act that seek to enhance the quality of education received by limited-English proficiency children whose native (primary) language is American Sign Language. (COED Recommendation #15)

Subsequently, Robert Davila, a deaf man who recently served as Assistant Secretary of the U.S. Department of Education, stated in a policy letter that "if a person is deaf or blind, or has no written language, the mode of communication would be that normally used by the person (such as sign language, Braille, or oral communication)" and "under EHA-B, testing or evaluation materials must be administered in a child's native language or other mode of communication appropriate to the child" (Davila, 1989, emphasis added).

These recommendations and subsequent responses make it very clear that assessment must be conducted in the deaf child's primary communication mode. Deaf children are very heterogeneous with respect to primary communication mode. For the benefit of readers who are unfamiliar with the variety of communication modes used by deaf children, the terms "ASL," various terms related to signed English, and "PSE" will be discussed briefly. COED has stated that ASL is a "full-fledged native minority language to which all of the provisions of the Bilingual Education Act should apply" (Commission on Education of the Deaf, 1988). ASL has its own grammatical rules (Wilbur, 1987). Many of these rules differ from the rules of English. Therefore, ASL signers cannot speak English and sign in ASL at the same time (Johnson, Liddell, & Erting, 1989).

Most manually coded forms of English (MCE) were developed in the 1970s by educators who were attempting to visually represent English because they hypothesized that speaking (English) and signing in English word order (as opposed to natural sign language/ASL word order) would help deaf students acquire English. These manually coded forms of English are called sign systems because they are no longer following the grammatical, morphological, phonological or lexical rules of ASL. All sign systems encourage the users to speak (English) and sign at the same time. However, each sign system has different rules about how to actually sign a sentence. In the case of most systems, these rules are written down; however, there is a great deal of variance between systems (Wilbur, 1987).

PSE, unlike the consciously-designed sign systems, develops wherever deaf and hearing people interact (Stokoe, 1990). It has also been called simultaneous communication or SimCom. Typically, it is a form of manual communication that makes use of ASL signs and grammatical features but primarily uses English word order. The degree to which PSE follows English word order varies with the user (e.g., hearing people tend to use more English conforming signs than do deaf people).

In summary, there is a fundamental concern on the part of much of the Deaf community and the professionals in the field of deafness that students who are deaf have significant communication mode barriers to overcome in order to have access to a free and appropriate public education that meets their unique educational needs (particularly their communication needs) (Commission on Education of the Deaf, 1988). The extensive recommendations made by COED to improve the quality of deaf education received "near-unanimity of the field in support of (their) report" (Bowe, 1991). Again, it is clear that traditional (paper-and-pencil) tests are still inappropriate because they are administered in written English only (without the visual sign accompanying them).

There are other deaf students who do not use any form of sign language and rely on an oral interpreter to translate English spoken words that are difficult to speech-read into a form that is easier for them to speech-read. In addition, oral interpreters are necessary if a particular teacher is difficult to speech-read. These students are expected to read English and take written English tests. However, for standardized tests, a teacher will, most likely, be giving the directions orally. Therefore, these students would require an oral interpreter to have "access" to assessments with oral directions.

Conditions for Inclusion/Exclusion of Students Who are Deaf from State and National Assessment

Standardized assessments are given to students for a variety of reasons. Purposes for assessment could include:

(a) Determination of optimal performance/student aptitude
(b) End-of-year achievement test
(c) Minimum standard assessments
(d) Data aggregation for making policy decisions
(e) Child count activities
(f) Placement in a special education program

This paper will be limited to addressing questions primarily related to the first three purposes listed above: aptitude, achievement and minimum standard assessments.

Aptitude assessments. If the purpose of the assessment is to determine a deaf child's aptitude or optimal performance, then it is imperative that the assessment not be confounded in any way. If the assessment is administered in oral or written standard English, the results will be confounded by the degree to which the student can speech-read and/or read and comprehend written English. As pointed out above, English may be a second language for a deaf student. In addition, if the student must write down answers (in English), the scoring of those answers will be confounded to the degree to which the student is a competent writer of English. In a study by Greenan and Loeding (1989), deaf students took a written assessment of their vocational skills which required them to write their own answers. These answers were to have been scored as indicating high, medium or low degree of skill, by individuals who were unfamiliar with deafness and by an expert in deafness. Naive scorers routinely agreed that the answers of the deaf students indicated low degree of skills. However, since the expert's ratings ranged from high to low, there was very poor reliability between the naive scorers and the expert in deafness, indicating that scoring of the written answers of deaf students could not be reliably performed by persons unfamiliar with deafness.

Achievement tests. If the purpose of the assessment is to determine how much a student has learned (i.e., achievement tests), then it is also imperative that the assessment not be confounded in any way. Therefore, just as with aptitude tests, achievement tests should be given in the student's primary communication mode. The only exception would be if the specific purpose of the assessment is to determine how much a deaf student has learned in the area of reading English or mastery of English composition skills. Then the test should be administered in written English, appropriate for the reading level of the student. Additional examples of assessments which should be given in English alone would be determination of the student's speech-reading abilities, and certain assessments given to determine the student's receptive and expressive, oral and written English language abilities.

Minimum standard assessments. If the purpose of the assessment is to determine whether the student should graduate, the assessment must be administered in the student's primary communication mode in order to be language-fair, according to both COED and the Department of Education. The issue of whether the test should be transliterated (directly translated into sign) or modified to reflect the content of the student's educational program remains. Restated, if students were in the special diploma track throughout high school, is it fair to give them the regular education track graduation test?

Proposed Guidelines for Determining Inclusion in Standardized Testing

Guidelines or criteria for determining who should be included in standardized testing or inclusion in an adapted assessment should require an examination of the student's primary communication mode, the student's use of an interpreter, and the student's functioning level. The following is a series of questions that may be used to examine these areas:

What is the student's primary communication mode?

(1) If it is not standard oral English (i.e., if it is American Sign Language, total communication, signed English, or another language), the student should not take the standardized assessment without adaptation for language difference. This viewpoint corresponds with the view of COED (1988) and the U.S. Department of Education (Davila, 1989).

(2) However, if the student's primary mode is standard oral English, the relevent questions to ask are as follows: Is the student expected to achieve at or above grade level by the IEP team? Does the student typically use an oral interpreter?

(a) If the student is expected to achieve at or above grade level and does not use an interpreter, then the student should take the standardized assessment without adaptation, unless there is an additional disability that would indicate the need for an adaptation. An example would be a deaf student who has a physical impairment that precludes that student from taking paper-pencil tests.

(b) If the student is expected to achieve at or above grade level and does use an oral interpreter, then the student should take the standardized assessment with an oral interpreter. We think it is preferable that the assessment be recorded with an oral interpreter on videodisc/CD-ROM. However, research needs to be conducted to determine whether there are significant differences between performance on traditional paper-pencil tests, tests using a live oral interpreter and tests using an oral interpreter presented with interactive videodisc/CD-ROM technology. It may be that oral deaf students will perform better when they have access to an on-site oral interpreter. However, multimedia technology should provide a more reliable, standardized way to provide that interpreting. In addition, the assessment should give the student the ability to re-view a question (just as normal students have the opportunity to re-read a test question). This is easily achieved with the use of multimedia assessments.

(c) If the student is expected to achieve below grade level, then the student should take a modified version of the assessment (unless the purpose of the assessment is to collect data on how all students perform on the assessment for data aggregation and policy decisions).

The Process of Decision Making

There are several options when deciding who meets the criteria for taking the standardized assessments:

(1) Exclude all students with hearing loss
(2) Include all students with hearing loss
(3) Include/exclude students based on degree of hearing loss
(4) Let teachers make decisions about who to include
(5) Let IEP teams make decisions about who to include

Options 1 and 2. The first two options (excluding or including all students with a hearing loss) would make decisions automatic. However, these two alternatives would run counter to the provisions of P.L. 94-142 and the Individuals with Disabilities Education Act (IDEA) which mandate that all students who are handicapped receive an appropriate education through the use of an individualized education plan (IEP) and be tested in their home or native language to guarantee nondiscriminatory testing. Certainly, determining what constitutes nondiscriminatory testing falls under the realm of an educational decision that must be made on a case-by-case basis. Inclusion of all students with a hearing loss, without translation of the assessment into sign language, would, in our opinion, be a mistake because this would force a great number of students to take assessments in their second language (English) and the results would only reflect their reading abilities in English and would not be valid measures of their true abilities in different subject areas. Moreover, inclusion of all students with a hearing loss in either the standardized assessment (written English, or standardized use of an oral interpreter or standardized language-fair sign version(s) of the assessment) or the modified assessment would be desirable.

Exclusion of all students with a hearing loss would be unwise. There are students with hearing loss who develop grade-appropriate ability to read and comprehend English. These students should compete with normal hearing students and take standardized assessments.

Option 3. Another alternative would be to make the decision to include or exclude a student based on hearing impairment level alone. Hearing impairment level alone is not sufficient to make this decision because it assumes a homogeneity of abilities that does not exist. The reading level of deaf students depends on a number of factors including development of necessary language and knowledge bases as well as the hearing status of the parents (King & Quigley, 1985; McGill-Franzen & Gormley, 1980; Gormley, 1981), socioeconomic status of the parents (Brasel & Quigley, 1977), age of onset and severity of hearing loss (Jensema, 1975), age at which student started school (Krupski, 1989), and quality of communication between student and the student's immediate family (Krupski, 1989). In addition, many deaf students may have learning disabilities, a physical impairment, or a vision impairment. Even deaf students with grade-appropriate reading levels still tend to have gaps between their spoken and written vocabularies that might adversely affect their performance on standardized assessments (McAnally, Rose, & Quigley, 1987).

Option 4. Another alternative might be to have the teacher of the deaf make decisions about who to include in standardized assessments, using the guidelines outlined above. Although the teacher of the deaf generally has a smaller caseload than a regular education teacher, and might know the individual students well, there are several problems with this alternative. Some deaf students are mainstreamed, with an interpreter, and do not see a teacher trained specifically for deaf students. Who would make the decision in these cases? Others only see a teacher of the deaf once a week on an itinerant basis. Yet other deaf students have a different teacher of the deaf for each academic area. Which of these teachers would make the decision? It appears that it would be extremely difficult to devise a set of guidelines from which teachers could make the inclusion/exclusion decision in every case. In addition, since the passage of P.L. 94-142, it has become accepted practice to rely on IEP teams to make educational decisions for students with hearing impairments severe enough to warrant special placement because this reduces the effect of personality conflicts and arbitrary decisions.

Option 5. The fifth alternative is to let IEP teams make decisions about who to include, using the guidelines outlined in the section above. This alternative appears to be the wisest way to make these decisions. The IEP team would have access to each student's cumulative records, previous IEPs and test scores. In this alternative, administrators, teachers and professionals providing related services, together with parents, and the students themselves, if they wish, would make the decision together. In this option, the parents or guardian would have the opportunity to approve or disapprove of the decision to include their child in such testing. For deaf students who are completely mainstreamed without support services, it seems reasonable that the IEP team would expect those students to take proposed state and national assessments without adaptation.

Kinds of Adaptations Permitted for Students Excluded from State and National Assessment

Kinds of adaptations currently used for deaf students will be discussed in the following section. The benefits and drawbacks of each adaptation will be explored. Currently, there are several kinds of alternative assessments for deaf students. These assessments consist of (a) giving students more time to complete tests; (b) administering paper and pencil tests that have been modified and standardized on deaf individuals (e.g., Stanford Achievement Test: Hearing Impaired version - SAT: HI); (c) direct transliteration (i.e., rendering oral English exactly, without omission) of tests into a manual sign system by interpreters (some are certified, others not) or teachers (most are not certified interpreters); (d) translation (interpretation) of tests into ASL by interpreters (some certified and others not) or teachers (most are not certified interpreters), (e) interactive videodisc-based assessments presented at modified reading levels with some version or versions of manual signs available (Bullis & Reiman, 1992; Educational Testing Service, 1991; Loeding & Crittenden, 1992, Reiman & Bullis, 1990). The advantages and disadvantages of each kind of adaptation will be discussed.

Additional time allotted. Giving students more time to complete tests is merited only if the students are slow readers who do comprehend what they read. In several studies, deaf students, who typically have below average English reading levels, did not demonstrate substantial benefit in their test scores from elimination of time limits (Conrad, 1979; Garrison & Coggiola, 1980). This type of adaptation would benefit a limited number of deaf students and therefore has restricted use in a proposed national assessment plan.

Paper and pencil tests. Paper and pencil tests have been modified and standardized for deaf individuals, such as the SAT:HI (see Allen, White, & Karchmer, 1983, and Rosenblaum, 1981, for guidelines on the modification process of materials for deaf students). Through modification, deaf students are able to demonstrate mastery (or lack of mastery) over the same information presented in the standardized test. Clarification of the instructions and additional examples are permissible, according to the administration guide (Center for Assessment and Demographic Studies, 1983, and subsequent revisions). The first advantage is that since the test has norms for deaf individuals, student results can be compared to peer groups. The second advantage is that following the initial modification and standardization costs, the tests cost no more than assessments for other individuals. In addition, no specialized equipment is needed. These tests still have the disadvantage of being administered in a language other than the primary language for many deaf students, which invalidates them for those students. In addition, the modification and standardization process needs to involve persons knowledgeable of Deaf Education.

In situ transliteration or interpretation of tests. Schools provide interpreters for students who are mainstreamed into regular education classes. It is conceivable that these interpreters could provide direct (word-for-word) transliteration of the proposed assessment into a manual English sign system or translation into ASL. However, the two main concerns to be addressed are:

(1) Which language or code should the test be rendered in (for each student)?

(2) Who is qualified to administer the test in the appropriate code/language?

The decision of what is the appropriate language or code needs to be made on a student-by-student basis by the IEP team. Often, students do not reliably determine the most appropriate version for themselves because they are not only unaware of the different types of sign language or systems which exist, they are unaware of what to call the communication mode that they use (Loeding & Crittenden, 1992).

Assuming transliteration of the assessment was deemed appropriate, transliteration would have the advantage of being potentially language-fair only for students who use SEE or signed English. A second advantage would be that if a qualified signer is used, the transliteration should yield a direct (word-for-word) translation each time. However, many school interpreters are not properly certified. The Registry of Interpreters for the Deaf is the certifying body for sign language interpreters in the United States. This organization has developed certification requirements for transliterating; however, separate certification requirements do not exist for each version of sign system nor do formally accepted standards for training or certifying educational interpreters exist in most states. It should be noted that many teachers of the deaf are not certified to interpret (let alone interpret in more than one version of manual signs). Ethically, people who are not qualified to interpret should not continue to interpret. Practically speaking, many administrators, teachers and some interpreters who work in educational settings reason that any interpreting is better than none. Realistically, this results in assessment situations where it is questionable whether the student truly received an appropriate direct transliteration of the assessment. Therefore, the reliability of this method of adaptation is suspect.

On the other hand, direct (word-for-word) transliteration is not possible or appropriate for students who communicate in ASL because ASL is a language distinct from English. For these students, a modified translation (interpretation of an assessment) is recommended. The advantage is that the assessment is potentially language-fair for students who use ASL. The Registry of Interpreters for the Deaf has established criteria for certifying ASL interpreters, but not specifically for educational interpreters. However, schools have the same difficulty finding qualified interpreters for ASL as they do for SEE 1, SEE 2 or Signed English.

To make the situation even more complicated, even certified interpreters would rarely sign a question in ASL a second or third time, using the identical signs, non manual features and facial expressions. This means that although live interpretation may increase the validity of assessment for certain students who use ASL, the test-retest reliability of such methods is highly questionable.

Multimedia-based assessments. Multimedia assessments that use videodisc, CD-ROM, CD-I, or Digital Video Interactive technology offer an innovative approach to provide a valid, reliable and equitable assessment for deaf individuals. To date, in the field of deafness, only videodisc-based assessments have been developed. One videodisc-based assessment for deaf students is a multiple-choice assessment designed for small group administration called the Transitional Competence Battery. The first version of this assessment uses only ASL while the second version uses only PSE (Bullis & Reiman, 1992). The test questions are presented at a modified reading and language complexity level. Data collected from over 230 deaf subjects, indicate acceptable levels of internal consistency reliability and test-retest reliability for five of the six subjects (above .75) (Bullis & Reiman, 1992). A videodisc-based assessment designed for individual use is currently at the prototype stage and has been developed for a portion of the Scholastic Aptitude Test (SAT). This prototype makes both ASL and English-order signs available (Educational Testing Service, 1991). Another assessment designed for individual administration is the Generalizable Interpersonal Vocational Skills Assessment for Hearing Impaired and Deaf Adolescents. This assessment is currently being tested to establish its test-retest reliability. This assessment makes both ASL and English-order signs available using a certified sign interpreter with English captions at modified reading and language complexity levels (Loeding & Crittenden, 1992). This particular assessment is administered and scored by the computer, with minimal assistance from the instructor.

If the proposed national assessment had a multimedia version, it would also include the capability to immediately score the student's performance and provide the instructor with a print-out of the results. Other advantages of the videodisc-based assessments with sign language available include being valid, reliable and much faster to administer and score than in-situ interpretation of assessments (Bullis & Reiman, 1992; Loeding & Crittenden, 1990, 1991, 1992; Educational Testing Service, 1991). The disadvantage is that not all schools presently have access to the necessary hardware to administer this type of assessment. However, this may be a temporary disadvantage. The state of Florida has provided funds so that every school in the state can purchase a videodisc player. Schools that have been designated "technology" schools also have the necessary hardware.

In summary, the cutting edge of assessment for deaf students will be multimedia-based assessments. These are the assessment practices of choice for students whose primary communication mode includes sign language or a manual sign code. Therefore, when the proposed state and national assessments are developed, every effort should be made to fund the development of matching interactive multimedia-based signed assessments at several reading levels in both ASL and English-order signs.

Analysis of the Extent to Which Adaptations Should be Permitted

This section of the paper will focus on providing a brief analysis of anticipated effects of videodisc-based adaptations on matters of validity, reliability and the equity of the results. Adaptations, such as presenting assessments using a certified interpreter recorded on a videodisc or compact disc, appear to offer the best option at this time to make the proposed assessment understandable to the greatest number of deaf students. These assessments may be able to bypass the variable of English reading ability that confounds the results of traditional assessment methods which are not language fair. Thus, deaf students will have access to a language-fair assessment and the validity of the proposed assessment will be increased.

Issues in development and standardization. Development and standardization of such multimedia-based assessment should make assessment more reliable. This technology ensures that the assessment will be signed the same way each time it is used. By taking videodisc or other multimedia-based assessments, deaf individuals will have valid, reliable, and equal access to the proposed state and national assessments.

Another question that needs to be addressed is what (English) reading level(s) should be used in preparation of standard assessments. A goal of Deaf Education should be achievement of at-grade-level-reading (and writing) proficiency for deaf students with normal cognition. However, reading levels of deaf children range from being illiterate to reading at the post-high school level. Most studies indicate that the average reading level of deaf high school seniors is roughly the equivalent of third or fourth grade (Allen, 1986). Therefore, it is important that at least one version of the assessment designed to test achievement (of any subject except Reading) be written at K-4th grade levels. More research is needed to determine whether the resulting sign language is at the K-4th grade level.

Assessment developers also need to address which combination of sign language and sign system should be made available in both (a) ASL and PSE or (b) ASL and SE or (c) ASL and SEE or (d) just one (ASL or PSE or SE or SEE)? Assuming that best practice indicates that deaf students need to take assessments in their primary communication mode, assessments need to be available in ASL, PSE and SE, with the ability to switch between signed versions when desired by the student. Recent research by Loeding and Crittenden (1992) indicates that when deaf adolescents have **both** ASL and PSE available to them, they elect to change signs occasionally to see how the question is signed in the other version of signs so that they were sure they understood the question before answering. Ideally, all assessments, including the proposed state and national assessments, should be available in whatever mode the students need.

The majority of deaf students being educated with some version of manual signs are being educated in either signed English or PSE (Bornstein, 1990). However, the number of students being educated in ASL is expected to grow because of the recommendations made in the document "Toward Equality" (Commission on Education of the Deaf, 1988) and the recommendations made in the document "Unlocking the Curriculum" (Johnson, 1989; Johnson, Lidell, & Erting, 1989). These authors are aware of a new school being started by Deaf individuals in Minnesota which has already received approval from a local school board to educate all of its students using ASL (Lange, 1992, personal communication), and a similar situation exists in Canada (Perigoe, 1993, personal communication). Very few students are being consistently educated in SEE. Research has shown that even in programs that say that they are using SEE, inconsistencies (teachers not being able to coordinate speaking and signing) have been noted (Bernstein, Maxwell, & Matthews, 1985, Kluwin, 1981; Marmor & Pettito, 1979). In order to improve the linguistic environment in such classrooms, both quantitatively and qualitatively, Stewart (1988) conducted research demonstrating the effectiveness of using two versions of sign (manually coded English and ASL); this procedure required that there be two teachers in each classroom. If assessments cannot be provided in all versions of sign, they should at least be provided in two versions to give students an additional chance to comprehend each question. ASL and PSE are being proposed as the two most reasonable versions of sign to use because it is expected that increasing numbers of deaf students will be using ASL. Research needs to be conducted with oral deaf students to see whether the PSE version (with the audio channel available, and an interpreter signing and speaking at the same time) is sufficient or if they need a separate version with the camera focusing nearer to the face and lips of the interpreter.

Decisions about modifications. In determining which modification is most appropriate for each hearing impaired student, the teacher's judgment should be relied upon. Ideally, the teacher would have a choice of reading level for the assessment and sign version(s) available to the student. With the advent of artificial intelligence, computers are capable of reliably making these decisions, as well, given the proper input. Loeding and Crittenden (1992) devised a screening program and a teacher questionnaire, both of which determined the primary communication mode for deaf adolescents. Data analysis is in progress to determine the extent to which the computer's decision matched the teacher's decision and the extent to which both predicted the mode used predominantly by each student in the subsequent assessment.

In view of the increased reliability and validity that videodisc-based systems offer, the National Symposium for Educational Applications of Technology for Deaf Students recently concluded that interactive, multimedia applications (such as interactive videodisc/CD-ROM-based assessment) clearly support positive educational outcomes for deaf students (Stuckless & Carroll, 1992). This organization forwarded the recommendation to the U.S. Department of Education to pursue research and development of interactive multimedia applications as a national priority for deaf students.

As our nation faces the future, we must strive to improve the education of all students and make assessments valid, reliable and accessible for all, including students with disabilities. For deaf students, this will involve preparing assessments in multiple versions of manual signs and a version appropriate for those who speech-read, using the innovative multimedia technology that is now available.

References

Allen, T. (1986). Patterns of academic achievement among hearing impaired students: 1974-1983. In A. Schildroth & M. Karchmer (Eds.), Deaf children in America (pp. 161-206). San Diego: College-Hill Press.

Allen, T., White, C., & Karchmer, M. (1983). Issues in the development of a special edition for hearing-impaired students of the 7th edition to the SAT. American Annals of the Deaf, 128, 34-39.

Bernstein, M., Maxwell, M., & Matthews, K. (1985). Bimodal and bilingual communication in schools for the deaf, Sign Language Studies, 47, 127-140.

Bornstein, H. (1990). Manual communication: Implications for education. Washington, DC: Gallendet University Press.

Bowe, F. (1991). Approaching equality: Education of the Deaf. Silver Spring, MD: T.J. Publishers.

Brasel, K., & Quigley, S. (1977). The influence of certain language and communication environments in early childhood on the development of language in deaf individuals.Journal of Speech and Hearing Research, 20, 95-107.

Bullis, M., & Reiman, J. (1992). Development and preliminary psychometric properties of the Transition Competence Battery for deaf adolescents and young adults. Exceptional Children, 59, 12-26.

Center for Assessment and Demographic Studies (1983). Administering the 1982 Stanford Achievement Test to hearing impaired students (7th ed.). Washington, DC: Gallaudet College, Center for Assessment and Demographic Studies.

Christensen, K. & Delgado, G. (1993). Multicultural issues in deafness. White Plains, NY: Longman.

Commission on Education of the Deaf (1988). Toward equality. A report to the President and the Congress of the United States. Washington, DC: U.S. Government Printing Office.

Conrad, R. (1979). The deaf school child. London: Harper & Row.

Davila, R. R. (1989). Letter to Mr. Robert Dawson, November 17 signed by Michael Vader, Acting Assistant Secretary of the U.S. Department of Education.

ERIC. (1989). Assessing the language difficulties of Hispanic bilingual students. Abstract 23. Reston, VA: ERIC Clearinghouse on Handicapped and Gifted Children.

Educational Testing Service. (1991). Sign language on the computer: Education and assessment. Princeton, NJ: Educational Testing Service.

Garrison, W., & Coggiola, D. (1980). Time limits in standardized testing: Effects on ability estimation (Paper Series No. 37). Rochester, NY: National Technical Institute for the Deaf.

Gormley, K. (1981). On the influence of familiarity on deaf students' text recall. American Annals of the Deaf, 126, 1024-1030.

Greenan, J., & Loeding, B. (1989). Inter-rater reliability and the Generalizable Interpersonal Vocational Skills Assessment. Working paper.

Jensema, C. (1975). The relationship between academic achievement and the demographic characteristics of hearing-impaired children and youth. Washington, D.C.: Gallaudet College, Office of Demographic Studies.

Johnson, R. C. (1989). Paper on use of ASL in teaching draws wide response at seminar. Research at Gallaudet. Washington, DC: Gallaudet Research Institute.

Johnson, R. E., Lidell, S. K., & Erting, C. J. (1989). *Unlocking the curriculum: Principles for achieving success in Deaf education* (GRI Working Paper 89-3) Washington, D.C.: Gallaudet Research Institute.

Jordan, I. K., & Karchmer, M. (1986). *Patterns of sign use among hearing impaired students*. In *Deaf children in America*, Schildroth, A. & Karchmer, M. (eds.) San Diego: College-Hill Press.

King, C. M., & Quigley, S. P. (1985). *Reading and Deafness*. San Diego: College-Hill Press.

Kluwin, T. (1981). The grammaticality of manual representations of English in classroom settings. *American Annals of the Deaf*, *127*, 417-421.

Krupski, A. (1989). *Encoding strategies used to process print information by prelingually, profoundly deaf children*. Doctoral dissertation. University of California, Los Angeles.

Loeding, B. L., & Crittenden, J. B. (1990). *The use of interactive videodisc-based assessment of interpersonal skills of youth who use sign language*. Proposal for funded U.S. Department of Education Grant Number G00180B00004.

Loeding, B. L., & Crittenden, J. B. (1991). *The use of interactive video-assisted assessment with youth who use sign language*. Proposal for funded U.S. Department of Education Grant Nanbar G00180B00004-91.

Loeding, B. L., & Crittenden, J. B. (1992). Presentation at the National Symposium for Educational Applications of Technology for Deaf Students held at the National Technical Institute for the Deaf, Rochester, New York (May, 1992).

Marmor, G. & Pettito, L. (1979). Simultaneous communication in the classroom: How well is English represented? *Sign Language Studies*, *23*, 99-136.

McAnally, P., Rose, S., & Quigley, S. (1987). *Language learning practices with Deaf children*. Boston: College-Hill.

McGill-Franzen, A., & Gormley, K. (1980). The influence of context on deaf readers' understanding of passive sentences. *American Annals of the Deaf*, 125, 937-942.

Reiman, J. & Bullis, M. (1990). *The transition competence battery for deaf adolescents and young adults*. Monmouth, OR: Teaching Research.

Stewart, D. (1988). *Implementing consistent linguistic input into total communication classrooms*. Proceedings of the 1988 OSEP Project Directors Meeting, Washington, DC.

Stokoe, W. C. (1990). A special issue on sign communication. *Sign Language Studies*, *69*, 291-294.

Stuckless, R. & Carroll, J. K. (1992). *National priorities pertaining to educational applications of technology for Deaf students*. National Symposium for Educational Applications of Technology for Deaf Students, Rochester, New York.

Wilbur, R. (1987). *American Sign Language: Linguistic and applied dimensions*. Boston, MA: Little, Brown.

Inclusion and Accommodation: "You can tell what is important to a society by the things it chooses to measure"

Jack Merwin
University of Minnesota

This quotation in the title seems appropriate in light of the decisions to selectively not include students with disabilities when doing state and national assessments.

It was requested that this paper address some general considerations, particularly technical considerations, regarding the exclusion of students with disabilities from state and national assessment efforts. It was further requested that it address five questions. The format to this response is organized to follow this pattern: general considerations followed by reactions to the questions posed. Two overriding considerations in addressing the general issue at hand relate to the implications for the interpretation (and potential misinterpretation) of the results and the implications for individuals who may or may not be excluded.

General Considerations

Sampling: Sampling is obviously a major consideration in the interpretation of results. A sample must be structured to support the intended use of results. For example, if the results are to be used for making decisions about individuals, sampling of the content of the assessment is critical for the reliability of the scores of those assessed and all members of the population must be included. If the goal is an index to reflect something about some defined group in specific states, or the nation, the sampling considerations are quite different.

The intended and potential unintended use of the results must be considered in designing the sample. No one to date has designed an assessment that addresses all of the concerns of all of the constituencies that hold interest in the results. The design must be responsive to the intended primary, and possibly principal secondary use of the results. If the results are primarily to arrive at indexes for states, or the nation, from the interpretation standpoint (not the excluded individual's standpoint) the inclusion or exclusion of groups with small base rates will make little difference in the aggregated results for this purpose.

The sampling must also anticipate as well as possible the agendas of groups who will want to use the results in ways other than the use intended by the assessors. Once such uses are anticipated, the design should purposefully either accommodate these uses by others with as technically sound a base as possible or make such uses totally indefensible. It would be naive to design the sample for a state or national assessment on the assumption that people will not attempt to disaggregate the results for their own purposes.

Staying with the matter of assessments for the purpose of obtaining indexes for large groups (i.e., states or nations), the ultimate use that is most defensible, and most probable, is comparisons. Content sampling to provide a technically defensible snapshot of performance of extended content at a single point in time is probably unrealistic. The intent to use the results for inter-unit vs. intra-unit comparisons is an aspect that must affect sampling (including the inclusion or exclusion of students with disabilities).

For inter-unit comparisons (e.g., states or school districts) differences in base rates of students with particular disabilities is likely to be a major political issue if all students are included - even if the sample is so large relative to the base rate for the students with disabilities, that their inclusion has little or no effect on the aggregated results. Intra-unit comparisons are used to show change in the aggregate index over time. Changes in the base rates of the students with disabilities over the interval will undoubtedly become a political basis for dismissing undesirable results. This will happen even if, again, the sample (whole or partial) is so large relative to the base rates of students with designated disabilities that their inclusion has little impact on the aggregated indexes.

If the aim of the assessment is to provide meaningful results for sub-groups, such interpretation must be accommodated by the sampling design. If the sub-groups are large, e.g., states or regions in a national assessment, such accommodation is relatively easy. If the sub-groups have small base rates, e.g., students with disabilities, or even students with a particular disability, there may be a need to over sample these groups in a stratified design as NAEP has done with inner-city students.

Validity: We can now turn to the issue of validity as it relates to individuals with disabilities and their inclusion or exclusion in state or national assessments. First, it would be unconscionable to force students to go through the motions of participation for political ends unless there is evidence to support a belief that the same constructs (or domains) are indeed being measured by the subgroups to be aggregated. The National Academy of Sciences' Panel on Testing of Handicapped People called for a four-year research effort by testing agencies to determine the validity and comparability of admissions tests administered to handicapped examinees. This included a study of results obtained under accommodations to various disabilities then in use in these national programs. The book Testing Handicapped People by Willingham et al. (1988) sets forth the findings of research by ETS in response to the edict of the Panel. Whether lack of validity or comparability comes from the disability or attempts to accommodate it in the assessment, it is this type of evidence that is needed to justify a heavy burden of participation by the students with disabilities. The one exception would be the personal feelings of the participants which might argue for inclusion in the assessment without inclusion in the aggregate results and without the above mentioned evidence in hand.

Reliability: Considerations of reliability must also relate to purpose. If use of the results is to reach the level of the individual, reliability of individual assessment becomes a major issue. With aggregated results, it is a lesser consideration. In fact, for large samples, matrix sampling to get aggregated results could lead to relatively low reliability for individual results (which could not be used) and lessen the burden of students with disabilities who could meaningfully be included.

Aggregation: The literature on aggregation of data should have a good deal to say relative to the matter of exclusion of students with disabilities from state and national assessments, though I have not had time to carefully examine it.

One study dramatically illustrates this potential. Jaeger (1992), in his invited address to Division D of the American Educational Research Association last spring examined the contribution of a number of subgroups to the variance in means of students from different countries. Based on Jaeger's analysis, it appears that some of the variables he found to contribute heavily to the variance in mean differences would surpass the variance contribution of differences in base rates of students with disabilities if all students physically able to participate (with needed accommodations) were included in an assessment. In Jaeger's analyses such variables as poverty rate for children in single-parent families, percent of children living in single-parent families, percent of children involved in divorce, and, percent of youth economically active made practical as well as statistically significant contributions to the variance in means. Further research on differences among states (in national assessments) and in regions and school districts within states might well support an argument that such subgroups contribute more to differences in results than subgroups of students with disabilities. For example, if an administrator is politically motivated to attempt to eliminate a subgroup that will lower a district's mean score, including students with disabilities and excluding those at the lower end of variables Jaeger found significant might be a better strategy.

Jaeger also found a beautiful example of what can happen in aggregating results. In a study he cites, the mean score of the total SAT test-taking population declined while the mean score of every major racial and ethnic group that composes the population of SAT test-takers increased. It is this type of anomaly that can occur due to difference in shapes and sizes of distributions of the subgroups that needs to be carefully examined when dealing with an argument that including students with disabilities will lower the group mean.

Practicality: I am sure that a number of decision makers regarding assessment design will oversimplify the situation in terms of the costs of development, administration and scoring of tests with adaptations needed for students with disabilities and the contribution of their scores to unit means. Such thinking ignores both the added complication of the definition of the population sampled and the element of fairness for students with disabilities who would like to participate.

Responses to Questions

Question 1. I cannot add to the answer to the first question of what alternative kinds of assessment now go on in the states that is given in NCEO's Technical Report 2. However, I cannot resist reacting to a couple aspects of that report. First, there appears to be so much subjectivity in selecting students for exclusion that implementation of guidelines probably varies widely from unit to unit for the same assessment. Second, differences in proclaimed rules for exclusion for different assessments makes comparisons and aggregation across them impossible in light of the issue at hand.

Question 2. Advantages and disadvantages of alternative ways that decisions might be made about inclusion are likely to vary with intended use of the results. If the results are to be used to make decisions about individuals (e.g., graduation or some sort of certification) everyone, with whatever technically defensible adaptation that may be necessary to complete the assessment, would have to be included. If sampling is to take place some students will be excluded on a random basis, including students with disabilities who are included in the population to be sampled. Politically minded administrators might conclude that excluding students with designated disabilities would 1) save money and, 2) increase mean scores for their units. The type of adaptation (e.g., large print vs. an amanuensis) and the number of each needed might justify the former. The latter, however, is questionable at best. While the administrator's conclusion might be wrong on the latter point, it also ignores both the implication for individuals excluded personally and the differences in degree of disability for students within a designated category.

A basic consideration is the extent to which the concerns of those excluded are to be taken into account in making the decision. At one extreme, if such concerns are to be given no consideration, economic and other considerations may argue for excluding relatively large numbers of students with disabilities. Such decisions would be made by designers of the assessment who would define the population to be sampled and administrators responsible for implementing the rules for exclusion. If such concerns are to be given primary focus, then more subjectivity regarding a particular individual is likely to enter in and credibility would be a factor. It seems that both the best interests of the student involved and the credibility would argue against an administrator or teacher (who might fear an evaluation based on the results) being the decision maker. It is probably best for the IEP team to make the decision if the best interest of the individual student is to be the focus of the decision.

Question 3: First, there will need to be a decision on the meaningfulness of participation for the results and for the student with the disability in terms of degree of the disability. For those who might meaningfully participate with adaptation, the element of validity, the degree to which there is evidence to believe that the construct (or domain) tested is the same with and without the adaptation, is a critical concern. While content validity and expert judgment would be needed at a minimum, hard evidence from studies with the instruments to be used would be the ultimate to be desired.

The expense of developing, administering and scoring a particular adaptation along with the number of students likely to be in the sample and using that adaptation will be a major consideration.

If the results are to be used for decisions about individuals (e.g., certification of some sort) the reliability of the individual's results, the difficulty of completing the task, and the desires of the student needing the adaptation should be given consideration. If the results are to be based on sampling (with or without matrix sampling which could reduce the burden of participation), reliability of the individual's results is of less concern and definition of the population to be sampled would determine the inclusion of students with or without disabilities. If students with disabilities needing adaptation to participate are summarily excluded, this would alter the definition of the population sampled and consequent interpretation of the results.

If sampling is by individuals (not classrooms, schools, etc. where everyone in the sample unit is tested) a student with a disability requiring an adaptation may be excluded simply by not falling in the sample. Negative feelings about exclusion should be ameliorated by the fact that others, with and without disabilities, in the person's class and school also do not fall in the sample. Desires to participate by individuals then would not be a basis for inclusion. If the sample units are classrooms or schools, as is often the case in state and national assessments, then exclusion of a student with a disability could be a "put down" in the eyes of that student.

Question 4: Intended use of the results should play a part in who makes the decision. As noted above, if the use is for decisions about individuals and aggregate results for comparing states, districts, or schools (suspected or real), it would be best for the IEP team to make the decision. Depending on the age and ability of the individual to participate, this should be done in consultation with that individual.

If the decision rests with a policy board at the national or state level, it should not be made without the involvement of personnel who work directly with students with disabilities who might be excluded and with experts in the field, particularly those who have worked specifically with difficulties in the testing of students with disabilities.

Question 5: The overriding criterion should be whether participation will involve a difficult and distasteful chore for the student involved. If so, that student should not be forced to participate regardless of whether the results are likely to be meaningful.

Someone must make a decision regarding the cost/benefits of the development and use of adaptations; costs including the effort required by the individual and benefits including the meaningfulness of participation for the individual.

Overall, definition of the population assessed and the sampling plan will ultimately settle the matter. This may well involve some cut-off in terms of degree of disability and consequent difficulty of participation. This in essence shifts the whole issue down one notch. If the population is defined as students in schools, districts, regions, or states of a given age or grade level, every student, regardless of disability, who can meaningfully participate with a valid adaptation should be included.

Some people are apparently bothered by the testing of students using necessary adaptations and excluding the results in aggregations. Under certain circumstances I am not. If the use of the results includes assisting in decisions about the individual, then that use and exclusion in the aggregate where there is doubt about the construct (or domain) tapped using the adaptation makes sense. While no student should be put through a personally difficult assessment where there is little hope for meaningful results, participation for the sake of personal desire not to be excluded might be justified even under these circumstances.

References

Jaeger, R.M. (1992). "World class" standards, choice, and privatization: Weak measurement serving presumptive policy. Vice-Presidential Address to Division D presented at the annual meeting of the American Educational Research Association, San Francisco.

Willingham, W.W., Ragosta, M., Bennett, R.E., Braun, H., Rock, D.A., & Powers, D.E. (Eds.). (1988). Testing handicapped people. Boston: Allyn & Bacon.

Consequences And Incentives: Implications For Inclusion/Exclusion Decisions Regarding Students With Disabilities In State And National Assessment Programs

Daniel J. Reschly
Iowa State University

The National Center on Educational Outcomes (NCEO) proposes a complex model of educational outcomes that incorporates six outcome domains, enabling variables and conditions, and educational resources, contexts, and inputs (NCEO, 1992b; Ysseldyke, Thurlow, Bruininks, Gilman, Deno, McGrew, & Shriner, 1992). The underlying values of the model reflect commitments to: (a) use of assessment information to guide policy decisions about educational resources and programs; (b) student outcomes as the most important criterion of educational effectiveness; and (c) inclusion of students with disabilities to the maximum extent possible in the assessment of outcomes. These value commitments are timely regarding the educational reform themes in the 1990s (Bruininks, Thurlow, & Ysseldyke, 1992) and the evolving, increasing commitment to inclusion to the fullest extent possible of persons with disabilities in normal settings and activities, with normal persons (NCEO, 1992a).

This paper is directed to one component of the broad concerns of NCEO with educational outcomes, specifically, the inclusion/exclusion of students with disabilities in state and national assessment programs. These large scale programs typically have the characteristics of: (a) assessment of literacy with less emphasis on other outcome domains; (b) use of group administered standardized measures; (c) Comparisons of groups such as classrooms or states rather than examination of individual progress; and (d) evaluation of educational programs with possible negative and positive consequences attached to high or low performance.

The NCEO staff compiled data suggesting that current practices regarding inclusion/exclusion in state and national assessment programs vary haphazardly among districts within states and between states (McGrew, Thurlow, Shriner, & Spiegel, 1992; NCEO, 1991, 1992c). Rules or suggestions have not been developed to guide these inclusion/exclusion decisions. Estimates compiled from state survey results suggest that as many as half of all students with disabilities may be excluded from state and national assessment programs. Many, perhaps most, of these exclusion decisions are unnecessary, with the undesirable consequences that further separation of disabled from non-disabled students is subtly reinforced, the effectiveness of special educational programs is not evaluated systematically, and policy makers are provided incomplete information on student performance. The issues regarding inclusion/exclusion decisions are discussed following some comments on the centrality of outcomes to long standing issues in special education.

Outcomes Criteria

The emphasis on outcomes is important for all students; for students with disabilities, it is especially critical. The use of an outcomes criterion was suggested as a means to evaluate the usefulness and fairness of assessment procedures and approaches used with minority students who were over-represented in special education (Reschly, 1980, 1988). Undocumented or non-significant effects have been the quintessential problem in justifying minority over-representation in special education and, more broadly, the additional costs of special education programs. Several meta-analyses of special education efficacy have suggested weak or undocumented effects on academic achievement criteria (see especially, Kavale, 1990). Other results suggest that application of the available knowledge base can produce more positive outcomes for students with disabilities

(Deno, 1985; Fuchs & Fuchs, 1986; Ysseldyke, 1984; Ysseldyke, Reynolds, & Weinberg, 1984). Systematic assessment of outcomes is essential to implementing state of the art principles in the education of students with disabilities.

Complexities Of Outcome Assessment

The NCEO focus on outcomes is consistent with national reform trends and addresses persistent issues in special education for students with disabilities. It seems absurd to not consider outcomes as the most important criteria. Although it is "intuitively obvious" that student outcomes should be our primary focus in assessing attainment of educational goals, implementation of the intuitively obvious usually is much more complex than it first appears.

This complexity is recognized by the NCEO in the specification of a model of educational outcomes that includes six different domains. The domains vary in content from those for which there are relatively well developed measures (literacy) to those where measures have limited or unknown technical adequacy (contribution/citizenship). The extent to which students with disabilities are included in state and national assessment of student outcomes in the six domains is further complicated by: (a) the **outcome domains** that are being assessed; (b) unresolved issues regarding the **purposes** of assessment and the **inferences** that will be made from the results of assessment programs; (c) **type** and **severity** of students' disabilities; and (d) the **measurement** procedures being used. Interactions among these four factors should be expected in real life situations involving assessment of outcomes.

Outcome Domains

The outcome domains represent diverse combinations of affective, academic, and social behaviors (NCEO, 1992b). The ease and appropriateness of assessment of outcomes for students with disabilities will vary by domain as well as other factors. For example, independence and responsibility are frequently assessed in post-school studies of the adjustment of students with disabilities, including all types and levels of disabilities. Satisfaction with special education programming and with current life circumstances also has been included in evaluations of special education programming during school and at post-school. These assessments may be conducted by asking rather direct questions to the person with disabilities, **OR** questioning a third party such as a parent. Assessment in these domains may now be more common with disabled than non-disabled students.

Domains such as contribution/citizenship, physical/mental health, and social/behavioral are not now, for the most part, assessed systematically with any students regardless of disability status. Systematic measurement in these areas will be increasingly feasible to the degree that consensus can be achieved on the critical indicators and effective measures of competence in these domains. I suspect that the indicators and measures in these domains will depend heavily on three sources of information: (a) unobtrusive, objective indices such as arrest records and participation in various community activities; (b) interview or survey with the individual; and (c) interview or survey with third parties. It would appear that these methods would be appropriate for nearly all students with disabilities.

The domain of literacy predominates in state or national assessment programs, and generally represents the most important educational outcome in the minds of the general citizenry. When comparisons are made of educational outcomes, be they cross-national, inter-state, inter-district, or inter-teacher, the target almost always is literacy measures involving language and mathematical competencies.

Literacy assessment is much more complex than appreciated by the general public or most policy makers. Establishing a national test taken by **ALL** students, with a variety of negative and positive consequences associated with poor or good performance, seems rational and feasible.

Opposition to a national uniform test and associated consequences often is interpreted as stemming from fuzzy-headed liberal views or vested interests in the educational status quo. There are, however, unresolved (and unresolvable) issues associated with attempts to delineate the specific literacy skills to be assessed (e. g., mathematical calculation vs. mathematical reasoning), the appropriate method of assessment (e. g., objective standardized tests vs. actual student products in portfolios), and methods to summarize results (e. g., comparisons to others with similar characteristics or comparisons to absolute standards). Not surprisingly, even more controversy emerges regarding explanations for high or low results and the appropriate consequences for low and high scoring educational units.

Literacy goals are emphasized in nearly all special education programs for students with mild disabilities. The results of three Iowa studies of educational programs for students with mild mental retardation (Reschly, Robinson, Volmer, & Wilson, 1988), specific learning disabilities (Kavale & Reece, 1992) and behavior disorders (Reiher & Reschly, 1990) (the most frequently occurring disabilities of all types with nearly all cases being mild in severity) indicated that reading and mathematics skills were the top two Individualized Educational Program (IEP) goals for students in each of these classifications. Mild disabilities constitute approximately 90% of all students with disabilities, or about 8% to 10% of all public school students (Algozzine & Korinek, 1985; Reschly, 1988; U.S. Department of Education, 1991)

Summary. Inclusion/exclusion decisions can be expected to vary by outcome domain. These variations have more to do with measurement processes than inappropriateness of the NCEO six domains for students with disabilities. Clearly, assessment of literacy with this population is appropriate; indeed, literacy goals dominate the mandated individualized programs for the vast majority of students with disabilities. The other five domains appear to be appropriate for virtually all students with disabilities. Current measurement operations for these five domains generally are appropriate for most students with disabilities. The critical problem is with the interaction of conventional measurement processes for literacy and the varying characteristics of students with disabilities.

Purposes and Inferences

A classic discussion of assessment purposes appears in Salvia and Ysseldyke (1988). Assessment purpose means the decision to be made using the information that is collected, while inference involves the relationship between the sample of behavior and the meaning attributed to that behavior. We may observe and record certain data during a sample of oral reading behavior. There could be several purposes for that activity; for now, let us assume the purpose is to assess individual student progress in reading. And different inferences might be made based on the observation and recording of oral reading behavior. These inferences may involve interpretations about various subskills or overall reading competence, motivation, effort, or (for some) interference from emotional conflicts. The point is that the behavior (oral reading) is the same, but the inference is different.

Understanding the significance, and then careful delineation, of purpose and inference is essential to any good assessment program and, perhaps, crucial to state and national assessment programs. My experience as a member of the State of Georgia Assessment Advisory Board (where state-wide educational assessment programs in Georgia and other states are reviewed) and the American Psychological Association Committee on Psychological Tests and Assessment (where various proposals for national literacy tests or assessment have been reviewed) indicates that assessment mechanics (e. g., item types), test content, and scaling typically dominate discussions at the expense of consideration of why the assessment is done, what will be assessed, what interpretation(s) will result, how the results will be used, and what consequences will be established for good and poor performance. Greater emphasis on the latter issues would both improve state and national assessment programs and clarify issues regarding inclusion/exclusion of students with disabilities.

Type and Severity of Disabilities

Two general types of disabilities are identified among school age populations, those for whom there **ARE, OR ARE NOT,** identifiable biological anomalies that are functionally related to behavioral limitations (Reschly, 1987). A medical model perspective best accounts for student disabilities with underlying biological anomalies while a social model is most appropriate when no underlying biological anomalies exist. The most common underlying biological anomalies are neurological, sensory, neuromuscular, and health. The categories of disability wherein there typically are biological anomalies are multiple disabilities, deafness, hard of hearing, orthopedic impairments, other health impairments, visual impairments, deaf-blindness, and moderate, severe, and profound levels of mental retardation. When most persons think of students with disabilities, they have in mind these students with biological impairments.

In fact, the vast majority of students with disabilities do not have identifiable biological anomalies that would interfere with participating in state and national assessment programs. Based on examining the national prevalence of the eleven categories of disability (U.S. Department of Education, 1991), I suspect that less than two per cent of the overall student population has a biological anomaly that would interfere with performance on assessment procedures like group administered standardized tests. I do not suggest that these students be ignored; their actual number needs to be recognized in order to establish a realistic view of the issues.

Level of severity varies significantly within the population of students with disabilities. Students with the most severe disabilities are not able to participate in any meaningful way in typical state and national assessment programs. Literacy skills of the kinds measured in these programs simply have not and, for some, never will be developed. Students with disabilities such as moderate, severe, or profound mental retardation, severe early infantile autism, multiple physical handicaps, or severe health conditions often do not have literacy skills close to the levels of students in the very early elementary grades. The literacy skills of many of these students may be functionally non-existent and impossible to measure with any current technology. Inclusion of extremely low performing students in state and national assessments of literacy would tell us nothing useful about these students or the programs they are in.

There is substantial overlap between disabilities with high severity and those where biological anomalies are present. The most severe disabilities almost always involve biological anomalies. The reverse, however, is not true! Students with biological anomalies perform at all levels of literacy, including the very top as well as at the very bottom of the normal distribution. There are, therefore, relatively few students with biological anomalies and very severe disabilities, less than 2% of the overall student population, or less than 15% of the population of students with disabilities.

The receptive or expressive language demands of traditional measures of literacy may interfere with the performance of students with disabilities involving sensory impairments, various physical handicaps, and diverse health impairments. For example, blind children obviously cannot read the items using the standard administration materials and procedures. For most deaf students, oral instructions will need to be delivered in augmented communication forms such as signing. Many students with neuromuscular deficits will not be able to fill in the small circles with a number two pencil as required by many current tests used in state and national assessment programs. And some students with health impairments may not be able to sustain attention and effort over the relatively large blocks of time required in many assessment programs. The students mentioned in these examples may have very good literacy skills; however, they cannot perform, or are significantly penalized by, the receptive or expressive language demands of the task. Suggestions regarding inclusion policies and accommodations for these students are provided in a later section.

The lengthy discussion of type and severity of disabilities should not cause us to lose sight of the fact that most disabilities are at the mild level with no evidence of biological anomaly. As noted by NCEO, reasonable estimates suggest that at least half of these students can participate without accommodations in state and national assessment programs. Although that is true, it must be acknowledged that some students with mild disabilities are penalized by their disability when performing on conventional literacy measures. The interference can take many forms. Students with poor reading skills may not be able to comprehend the instructions on a test or the elements of a written mathematics "thought problem" even though they have mastered the competencies measured by the test (e. g., mathematical reasoning). Students with high levels of impulsiveness, distractibility, and hyperactivity are unlikely to be able to concentrate sufficiently on conventional group administered measures to perform to the actual level of their literacy skills. Finally, the skills of certain students in specific academic areas such as reading may be so low that the reading demands of the test are impossible for them to perform. The student may then earn a very low, but invalid score. If the purpose of the test is to measure reading, then the low score likely is a valid indicator of the student's reading competence. If, however, the purpose of the test was to assess intellectual mastery of citizenship rights and obligations, the student's low score is invalid because the receptive language demands of the test prevented the student from exhibiting what s/he knew in that domain. Many other examples of contamination of literacy measure results by non-related task demands could be cited. The critical issue is careful matching of measurement operations and task demands to assessment content and purpose.

Summary. The appropriateness of inclusion/exclusion of students with disabilities interacts with level of severity and type of disability. If disability severity and type are not properly taken into account, assessment results for as many as half of all students with disabilities may reflect the effects of the disability rather than the competency being assessed. Severity and type are most likely to influence literacy assessment, the principal focus of most current state and national assessment programs. Inclusion of these students without accommodations would satisfy certain goals. The costs, however, are likely to be enormous. The assessment results for those individuals are unlikely to be meaningful or valid and, perhaps most important, the credibility and acceptability of state and national assessment programs will be diminished. Persons affected by the results may then see them as unfair and arbitrary, especially those who experience negative sanctions (see later discussion of consequences or stakes). Credibility and acceptability, although beyond the scope of this paper, are crucial to the success of state and national assessment.

Measurement Procedures

Knoff and Batsch (1991) pointed out that there are four ways to collect assessment information: **test** the student directly, **interview** the student or other persons, **review** records, and **observe** student behavior. The latter three methods, with the exception of interviewing the student, are likely to be as appropriate for disabled as non-disabled as long as the behaviors are within the repertoire of the student. The latter three methods are more likely to be used with affective and social domains of behavior while testing is the most common method of assessing literacy in state and national assessment programs.

The appropriateness of inclusion/exclusion of students with disabilities in state and national testing programs is influenced by the interaction of different methods of testing students' competencies with the other factors discussed above (outcome domain, purpose and inference, and level and severity of disability). Different testing methods place different demands on students' receptive and expressive language competencies. Group administered literacy measures, the most frequently used method in state and national assessment programs, place extensive demands on receptive and expressive language in order to understand and respond to the test items. Some examples of the required skills that may or may not be related to what the test items are supposed to measure are: (a) hearing and comprehension of the spoken word; (b) sufficient sight to see items and enter answers; (c) adequate reading skills in order to follow instructions and to comprehend the items; and (d) adequate psychomotor skills to mark answers. Deficiencies in any of these skills likely would produce lower scores than appropriate in view of the student's actual competencies. A seemingly simple solution is to change the administration procedures to minimize the disability through such accommodations as reading instructions, allowing someone else to enter answers that are selected orally by the student, and so on. The problem with these seemingly innocuous accommodations is that they do not conform to standardization procedures on which the test norms and other interpretative devices are based (see later discussion of accommodations).

Individually administered literacy measures typically make fewer demands on receptive and expressive language skills. Such tests usually require little or no reading, and greater flexibility often exists regarding communicating the instructions. The individually administered measures are not, however, free of receptive or expressive language demands. For example, directions often are read orally, requiring that the student be able to comprehend the spoken word. Individually administered literacy tests also may require psychomotor skills such as writing words or digits and pointing to response choices. Accommodations for some students with disabilities are needed even on the individually administered tests. Individually administered are much more expensive than group administered measures, a not insignificant consideration in decisions about state and national assessment programs.

Assessment Consequences and Inclusion/Exclusion

State and national assessment programs have differing consequences. The magnitude and importance of the consequences associated with assessment recently has been called "stakes," with high stakes assessment receiving increasing attention. High stakes assessment generally means that the results will be used to make important decisions that potentially have significant effects on the lives of students and educational professionals. The general trend appears to be increasing the stakes associated with state and national assessment programs.

The consequences or stakes associated with state or national assessment of outcomes vary on several dimensions. Three important dimensions are: (a) the size of the consequences; (b) the valence (positive or negative) of the consequence; and (c) the recipient of the consequence. Loss of job for a classroom teacher because students performed poorly on a literacy test would be a large negative consequence experienced by an individual professional. Loss of school district independence through the state department of education assuming control of the district would be a large negative consequence experienced by the community that had varying implications for individuals (consider the likely fate of the superintendent). Recent reports by the United States Secretary of Education identifying the states with the highest achievement scores (notoriety shared by Iowa and Wisconsin) is a positive consequence of national assessment that now has relatively small consequences that are experienced in a rather diffuse manner by many persons.

Some likely negative features of state and national assessment programs with large consequences are: (a) Strong pressure on educational professionals and students to produce positive results; (b) Perceptions that programs are zero sum games in which there are an equal number of winners and losers; (c) Concerns that existing background characteristics of students and communities create unfair competition that discriminates against some students and educational professionals; and (d) Perceptions that standardized content and administration procedures are unfair to many students. The negative perceptions create conditions ripe for outright cheating and unwarranted exclusion of students with disabilities or low achievement.

The factors of the magnitude, valence, and recipient(s) of consequences have vast implications for whether students with disabilities are included or excluded from state and national testing programs. There is an obvious general principle. The higher the stakes, the more likely is the unwarranted exclusion of low achieving students. The unfortunate corollary to this general principle is that as the stakes become higher, cheating also becomes more likely.

Unwarranted exclusion. The NCEO has collected survey data from State Directors of Special Education regarding the participation of students with disabilities in state and national testing programs. The results are disturbing. State policies vary widely and decisions appear to be haphazard and capricious (McGrew et al., 1992). Since most states do not have policies on participation, enormous variation almost certainly exists between districts within states. The source of this information also should be noted; state directors or their designees may have little evidence on which to base their opinions of district practices regarding inclusion/exclusion. The NCEO probably should gather additional information of actual practices at the local level. In fact, we have little systematic evidence on what actually is done at the local level. Anecdotal evidence suggests some serious abuses that may be widespread.

Unwarranted exclusion can be defined as directed or arranged non-participation in state or national assessment programs involving students for whom the assessment content is appropriate to curriculum goals pursued in their educational programs and the receptive or expressive language demands of the assessment tasks are within the student's behavioral repertoire. The motivation for unwarranted exclusion is obvious; excluding lower performing students enhances the average levels of performance in classrooms, school buildings, districts, and states. And most students with disabilities do perform at significantly below average levels.

Examples of appropriate exclusion include non-participation of: (a) blind students in group-administered standardized achievement tests requiring use of sight; (b) profoundly retarded students for whom literacy goals are either non-existent or are far below the skill levels of average kindergarten students; and (c) physically handicapped students who cannot perform the required responses or for whom an inordinate amount of time is required to make the responses.

There are many more students with disabilities who may be inappropriately excluded simply because their performance is expected to lower classroom, district, or state averages. Some examples of inappropriate exclusion include the directed or arranged non-participation of: (a) all students with disabilities; (b) students with IEP reading goals in standardized reading tests; and (c) students in Chapter I mathematics programs in standardized mathematics tests.

The methods of exclusion vary. In some cases there is straightforward directives to principals and teachers to not include any special education participants in standardized testing. Other methods are more subtle and difficult to document. Some anecdotes that have been communicated to me include encouraging certain students to stay home on days when standardized achievement tests are given, counting certain students as absent (on the test report records) even though the students were in school, and indicating that the answer sheet was not completed properly.

Cheating. Some of the examples of unwarranted exclusion discussed above are close to, if not beyond, the boundaries of ethical principles. Other practices related to state and national assessment programs can only be characterized accurately as cheating. Many of these practices are egregious violations of ethical and, in many cases, legal provisions.

Cheating by teaching test content and excluding low achieving students is at least suggested by the now well established and widely publicized Lake Woebegone Effect ("all the children are above average") in which virtually every state in the late 1980s managed to obtain scores on standardized achievement tests that were above the national median. This truly miraculous accomplishment suggests that standardized test results can be significantly influenced by factors other than actual student competence, and that educational professionals will cheat under certain conditions.

The conditions under which human beings cheat have been relatively well known from decades of research on the phenomenon. Most persons will cheat, including educational professionals, when the following variables are beyond moderate levels: (a) pressure to perform, i.e., there are significant negative consequences; (b) beliefs that no good alternatives to cheating are available to attain success or avoid failure; (c) opportunity to cheat with small likelihood of getting caught; and (d) perceptions that the evaluation process is unfair. Many teachers and other educational professionals do, indeed, regard their circumstances in relation to state and national assessment programs with high consequences as conforming to these four conditions. And as the stakes become higher, with such tangible consequences as salary and job security, we should expect more problems with unwarranted exclusion of students with disabilities and outright cheating.

Inclusion/Exclusion Policy Alternatives

The policy alternatives suggested here are specific to the assessment purposes and contexts that are the focus of this paper, that is, state and national assessment programs that are likely to involve high stakes decisions. For clarity, it is important to note that the National Center on Educational Outcomes (NCEO) has a broader mission than suggested by the focus of this paper. The author as well as the NCEO staff recognize multiple assessment purposes and contexts and share commitments to emphasizing outcomes in multiple domains for students with and without disabilities. The concern here is inappropriate exclusion of students with disabilities from participation in state and national assessment programs that, currently, are likely to examine the acquisition of literacy skills using standardized assessment procedures. A variety of policy alternatives are considered in the following sections.

Exclude All Students with Disabilities

The proportion of students with disabilities placed in special education varies by a relatively small amount among the states. Virtually all states identify from 9.5% to 12% of their students as educationally disabled and in need of special education. One policy alternative is to simply exclude all students with disabilities who are in special education programs. Presence of an IEP and legal status as a student with a disability could be used to operationalize this policy. As long as the sampling units (states or districts) make the exclusion decisions on the same bases, and as long as relatively equal proportions of students in the units being compared are classified as disabled, accurate comparisons should result of levels of student literacy.

Several advantages of exclusion of all students with disabilities from state and national assessment of literacy can be identified. First, local educators are likely to see this solution as fair and appropriate. It is easy and straightforward to apply and can be implemented consistently by educational professionals throughout the nation. Some consideration might be made for the slight variations among states in prevalence of students with disabilities served by special education, but the score adjustments would not need to be large and there would be relatively little effect on the relative standing of states.

This approach, however, has significant disadvantages. Among them are the wide variations among districts and buildings in overall proportion of students classified as disabled and placed in special education; building or district comparisons could not be made using this method. Another significant disadvantage is the reinforcement of the traditional exclusion of persons with disabilities from normal activities. Finally, scores could not be reported involving descriptions of skills or competencies in the general population of students since about 10% to 12% of students would be excluded. Statements such as the following could not be made, "50% of all tenth grade students in Iowa and Minnesota can apply chaos theory to fluctuations in the price of hogs at interior markets." In other words, all score reporting citing skills or competencies would have to be constrained to non-disabled students (not the general population of students) or be restricted to normative comparisons by district or state (e.g., status standard scores or percentiles).

Include **ALL** Students with Disabilities in Score Reporting

This policy may seem absurd at first glance. Participation of all students with disabilities in score reporting, with no exceptions, could be required. Those students with disabilities that were of a type or level of severity that actual participation was impossible could be assigned the lowest standard score or a percentile rank of one. The score assignment procedures for these students would have to be standardized across districts or states, a relatively easy operation.

The major advantage of this policy alternative is that clear incentives would exist for local officials to foster participation by students with disabilities, since those who did participate could not, by definition, be assigned the lowest possible score. Greater proportions of students with disabilities would participate in the state and national assessment programs, accurate comparisons of educational units could be made, and accurate proportions of students mastering various competencies could be reported.

Some disadvantages are obvious. Some students for whom participation would be extremely unpleasant might be forced to attempt to complete standardized examinations that have little or no direct or indirect benefit to them. This policy might be perceived as unfair and even cruel. Finally, the policy would foster participation that simply was not meaningful to some examinees or to educational policy makers.

Some variations of total exclusion or total inclusion are conceivable. For example, each district might be allowed to exclude 2% of its population because that proportion of students have severe disabilities that preclude meaningful participation. All other students would have to participate with low scores automatically recorded for non-participants. This slightly more moderate policy likely would be more acceptable and avoid the worst of the cases of non-meaningful participation by students with disabilities.

Incentives are essential to foster greater participation of students with disabilities in state and national assessment programs. These incentives should include reinforcement and response cost contingencies. Score aggregation and summarization could reflect these incentives with exclusion negatively sanctioned by very low score assignments to students who did not participate.

Accommodations

Participating in state and national assessment programs could be increased slightly by allowing variations in standardized procedures to accommodate deficits in receptive and expressive language. A small proportion of students with disabilities has sensory impairments such as hearing or visual losses or neuromotor problems. Some of these students have sufficiently developed literacy competencies to permit participation in standardized assessment procedures **IF** changes were made to accommodate their disabilities. These changes might involve reading items to students, presentation of the items through Braille, and signing test directions.

The accommodations would increase participation, a desirable goal. At the same time the accommodations would complicate interpretation of individual scores; for example, are comparisons to norms for the general population appropriate for a blind student for whom the items were read?

These accommodations, however, would not necessarily bias interpretation of results for districts or states. First, standard guidelines could be established for making accommodations. As long as all districts and states followed the guidelines, district or state comparisons would not be compromised. Furthermore, a small proportion of students would be affected. I suspect that the overall influence would be very small on district or state comparisons even if the accommodation guidelines were not followed with high consistency. In short, accommodations through changing receptive or expressive language demands or response mode will have a trivial effect on district or state comparisons. Establishing relatively liberal policies regarding such accommodations, communicating those policies and encouraging their implementation by all districts and states, likely would increase the perception of fairness and the credibility of the assessment program.

Enhancing Integrity

The loss of integrity may be the greatest threat to the success of state and national assessment programs. Unethical practices or cheating are likely in high stakes assessment programs under certain conditions (see prior discussion). One of the major ways that integrity is diminished is through unwarranted exclusion of students with disabilities and violations of test security.

The following steps are likely to preserve and enhance the integrity of state and national testing programs. First, educational professionals, parents, and students should be provided thorough descriptions of expected competencies well before the implementation of high stakes assessment. These competencies should involve generalizable academic skills and cognitive operations. If the competencies are formulated properly, we should hope that teachers "teach to the test." Clear descriptions of objectives and sample items should accompany the descriptions of skills and competencies. These steps increase fairness, both as a perception and as a reality. Persons are more likely to protect the integrity of a process that they view as fair.

Second, the **use** of the results of state and national assessment programs should not be restricted to high stakes comparisons of classrooms, buildings, districts, and states. In addition to these comparisons, feedback to teachers and students should be provided with top priority given to improving skills and competencies. If such feedback is provided, teachers and students are more likely to have attributions related to time and effort rather than seeing successful performance on the assessment process as being beyond legitimate efforts. Such perceptions are important to enhancing the integrity of assessment programs.

Third, test security needs to be rigorously protected. For example, persons other than those to be held accountable by student performance might be assigned to proctor administration of assessment procedures. Some of the decisions about whether students with disabilities participate and the degree to which standardized procedures are followed probably should be taken out of the hands of those most directly affected by the results.

Fourth, consciousness should be raised about unethical or illegal practices. Some educational professionals simply do not know what is appropriate regarding assessment practices; others need to be reminded. This step alone would, I suspect, enhance integrity to a significant degree.

Fifth, the zero sum, winners and losers, features of high stakes assessment programs should be minimized to the greatest extent possible. Win-Win situations need to be created, with positive incentives established that are within the reach of all participants. Failure to establish realistic positive incentives almost ensures that integrity will be a problem as the stakes become higher!

Integrity is a significant issue that influences the inclusion of students with disabilities. There are reasonable steps that can improve integrity; fortunately, these steps are educationally sound and consistent with the interests of all constituencies for state and national assessment programs.

Summary

This paper focused on one of the many goals established by the NCEO, specifically, ways to increase the participation of students with disabilities in state and national assessment programs. Increased participation is desirable for many reasons. The typical situation addressed in this paper involves assessment of literacy with group-administered standardized tests. Reasons for non-participation of students with disabilities were analyzed, and suggestions made for fostering greater participation. In addition, this paper emphasized accommodations in testing, procedures to reward or punish non-participation, and methods to improve the overall integrity of assessment programs. The majority of cases of unwarranted exclusion may be, in the end, a matter of integrity.

References

Algozzine, B, & Korinek, L. (1985). Where is special education for students with high prevalence handicaps going? Exceptional Children, 51, 388-394.

Bruininks, R., Thurlow, M. L., & Ysseldyke, J. E. (1992). Assessing the right outcomes:Prospects for improving education for youth with disabilities. Education and Training in Mental Retardation, 27, 93-100.

Deno, S. L. (1985). Curriculum-based assessment: The emerging alternative. Exceptional Children, 52, 219-232.

Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. Exceptional Children, 53, 199-208.

Kavale, K. (1990). The effectiveness of special education. In T. B. Gutkin & C. R. Reynolds (Eds.), The handbook of school psychology (2nd ed.) (pp. 868-898). New York: Wiley.

Kavale, K. A., & Reece, J. H. (1992). The character of learning disabilities: An Iowa profile. Learning Disability Quarterly, 15, 74-94.

Knoff, H. M., & Batsche, G. M. (1991). Integrating school and educational psychology to meet the educational and mental health needs of all children. Educational Psychologist, 26,167-183.

McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. N. (1992). Inclusion of students with disabilities in national and state data collection programs. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

NCEO. (1991). Students excluded from education data. Outcomes, 1 (2), 1.

NCEO. (1992a). Including students with disabilities in national andstate data collection programs (Brief Report 1). Minneapolis, MN: National Center on Educational Outcomes.

NCEO. (1992b). NCEO revises model of educational outcomes. Outcomes, 2 (1), 5.

NCEO. (1992c). Results of NCEO survey appear in report. Outcomes, 1 (2), 2.

Reiher, T. C., & Reschly, D. J. (1990). Iowa Behavior Disorders Research Project: Final Report. Des Moines, IA: Iowa Department of Education, Bureau of Special Education.

Reschly, D. (1980). School psychologists and assessment in the future. Professional Psychology: Research and Practice, 11, 841-848.

Reschly, D. J. (1987). Learning characteristics of mildly handicapped children: Implications for classification, placement, and programming. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), The handbook of special education: Research and practice (Vol.I) (pp. 35-58). Oxford, England: Pergamon Press.

Reschly, D. J. (1988). Special education reform: School psychology revolution. School Psychology Review, 17, 459-475.

Reschly, D. J., Robinson, G. A., Volmer, L. M., & Wilson, L. R. (1988). Iowa mental disabilities research project final report. Des Moines, IA: Iowa Department of Education, Bureau of Special Education.

Salvia, J., & Ysseldyke, J. E. (1988). Assessment in special and remedial education (4th Ed.). Boston: Houghton-Mifflin.

U.S. Department of Education. (1991). Thirteenth annual report to Congress on the implementation of the Individuals with Disabilities Act. Washington, DC: Author.

Ysseldyke, J. E. (Ed.) (1984). School psychology: The state of the art. Minneapolis, MN: University of Minnesota, National School Psychology Inservice Training Network.

Ysseldyke, J. E., Reynolds, M. C., & Weinberg, R. A. (1984). School psychology: A blueprint for training and practice. Minneapolis, MN: University of Minnesota, National School Psychology Inservice Training Network.

Ysseldyke, J. E., Thurlow, M. L., Bruininks, R. H., Gilman, C. J., Deno, S. L., McGrew, K. S., & Shriner, J. G. (1992). An evolving conceptual model of educational outcomes for children and youth with disabilities (Working Paper 2). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Inclusion and Accommodation in Assessment at the Margins

Maynard C. Reynolds
University of Minnesota

Many voices -- including even those of the President and many state governors -- are heard these days advocating raised standards for learning by all students and calling for improved assessments to reflect how well students are meeting these standards (National Council on Education Standards and Testing, 1992). But there is much uncertainty about who and what should be involved. It is in this busy but ambiguous context that this critique of present practices in assessment and a set of proposals, hopefully constructive ones, are presented. A particular focus here is on "disabled" students or those who, for whatever reason, are marginal in their school achievements and behavior.

Unfortunately, assessment practices -- like bowling balls -- are often directed mainly "down the middle." Low achieving and disabled students are often excluded even in test standardization processes as well as in regular school assessment programs (McGrew, Thurlow, Shriner, & Spiegel, 1992). Such exclusions cause inflationary and misleading general descriptions of student learning. It's a way to make the general outcomes of schooling look better than they really are. Setting marginal students aside at testing time is also a part of a larger pattern of neglects and exclusions in the schools. The argument here will be to favor universal assessment practices, but in carefully specified domains.

It may have been acceptable in earlier years to pay but little attention to marginal students, those who were misbehaving or struggling for survival in their academic work. Assessment was one step toward removal of such students from the schools. Ralph Tyler (1987) tells us that "as late as 1900, 61 percent of the United States labor force were employed in unskilled jobs. . . Today, less than 5 percent . . . are engaged in . . . unskilled work" (p. viii). "In 1910, more than half the children had dropped out before completing sixth grade" (Tyler, 1987, p. ix). But, in these days there is a steep price to be paid, in social and economic terms, when one's education is neglected anywhere, even at the margins. Educating all children and youth, including those most inconvenient and difficult to serve, means facing up to truly complex problems. These remarks are intended to help chart a way through the assessment aspects of this difficult territory.

In considering the assessment aspect of a commitment to all students, one is mindful of a general failure by educators and psychologists to communicate to the general public and to policy leaders the facts of individual differences among students. It should be clear, but apparently it is not, that an achievement standard that all (or nearly all) students meet is a minimal standard and not a "world class" standard. That is the case in mathematics and other school subjects as truly as in "world class" swimming, chess, or operatic singing. Equally, it should be clear, but apparently is not, that good teaching improves learning rates, but that it does not produce uniformity among students in achievement levels. It should be clear, but may not be, that dichotomous decision rules based on achievement testing, such as deciding on promotion vs. retention in grade placements, or granting vs. denying high school diplomas, are grossly neglectful of mostly continuous differences in learning rates and likely to be unfair to racially minority students. Establishing the validity and fairness of such rules is a large and very complicated challenge, going well beyond questions of test validity. Perhaps some of these problems relate simply to extravagance in word usage, such as "world class," but clearly there are many misunderstandings to be reformed as one considers ideas and practices in assessment, standard setting, and instruction.

Some Assumptions and Beliefs

The following brief statements of belief and assumptions will provide an orienting preface to the major propositions of this paper.

- Assessment practices should be an integral part of efforts to enhance the instruction and learning of children and youth; otherwise they will be subverted and ultimately be eliminated.

- Assessment practices should meet high standards of reliability and validity; this applies not only to tests and other measuring tools, but also to broader matters of assessment and to decisions based on assessments. There is much talk these days about authentic assessment, portfolios as representative of achievements and like matters, some of it quite void of technical considerations. It seems timely to pay attention to certain technical standards for all assessment practices and related decisions. This is a particularly critical matter for students at the margins because they are disproportionately the ones about whom major decisions, sometimes of catastrophic import, are made on the basis of assessment procedures.

- Assessment practices that lead to classification, labeling, and special school placements of students are justified only when distinctly "special" instructional procedures are indicated for selected students and when there is evidence that using such "special" procedures will be enhancing in the lives and learning of the selected students. Colossal errors have been made in American schools in the overuse of assessment procedures to classify, label, and isolate so-called "disabled" and at-risk students. A special panel created by the National Academy of Sciences (NAS) to study placement practices in special education, especially in the case of so-called educable mentally retarded (EMR) students, put the standard this way: "In the context of educational decision making it is not enough to know that IQ tests predict future classroom performance, nor would it be enough even to know that they measure general ability. It is necessary to ask whether IQ tests provide information that leads to more effective instruction than would otherwise be possible" (Heller, Holtzman, & Messick, 1982, p. 53): "What is needed is evidence that children with scores in the EMR range will learn more effectively in a special program or placement" (p. 61). Millions of children have been separated from regular school classes in favor of isolation in special schools and classes by assessment procedures and decision rules that fall far short of standards suggested by the NAS panel. Indeed, it appears that a favorite way of dealing with children formerly excluded from schools is now to admit them but to isolate them in categorical programs.

- Sinclair and Ghory (1987) are correct, I assume, in their assertion that "the same pedagogic strategies and environmental adjustments needed to help teachers relate more effectively with marginal learners are also promising approaches for adaptive and productive instruction that will benefit all learners" (p. 159). In a meta-review of research literature, followed by a survey of teachers, Reynolds, Wang, and Walberg (1992) found that the same principles of instruction were judged to be important by teachers in general education and in special education. This suggests that some of the separations that have developed in teacher preparation between general and special education are unjustified and that more can and should be done in integrated programs.

- In the cases of students who show very low rates of learning, it is important to study the student, his/her life situation, and the instructional program used in the period of failure. This is to say that assessments should run to both student and programmatic targets with a view toward designing a more promising environment and instructional program. In framing such a dual strategy for assessment we should be guided by the well-confirmed evidence about what has effects upon student learning as well as to learning outcomes themselves. The strategy should give emphasis to variables that can be manipulated by teachers and parents rather than to those that are static. The"bottom line," of course, is the student outcomes. All is for naught except as learning outcomes are improved.

- It appears that there is a trend in the schools to prefer domain-referenced (meant here to be synonymous with objectives-referenced or syllabus-referenced) assessment procedures over norm-referenced procedures (Hively & Reynolds, 1975). This proposes a tightened linkage between what is taught (the curriculum) and what is tested. It reduces emphasis on the psychometric properties of items used in tests in favor of more emphasis on fidelity to curriculum. Mere prediction of academic progress is reduced in value in favor of assessment procedures that have usefulness in managing instruction. This trend is regarded favorably here, but with acknowledgment that this kind of changeover involves many technical and attitudinal difficulties (Koretz, Madus, Haertel, & Benton, 1992).

Universal Assessment

Under what conditions should assessments be made of literally all students? In considering this question it is useful to make a distinction between subject matters and skills that are required of all students (the cultural imperatives) and those that are deemed to be not so uniformly essential (the cultural electives). The terms cultural imperatives and cultural electives were first used, as far as this author is aware, by George Stoddard in his book The Dual Progress Plan (1961). In a complex culture there are some areas of learning that are truly essential for thriving as human beings. These are the imperatives, the basics of education, because they are the tools of the culture without which the individual is confined by fundamental ignorance and limited choice.

In the United States in the final years of the 20th century the cultural imperatives, I think, are these:

1. Language--in all receptive and expressive forms, including speaking, listening, comprehending, reading, and writing.
2. Mathematics--including counting, measuring, and computing.
3. Social Skills--abilities involved in cooperation, group life, and good citizenship in a democratic society; also avoidance of destructive behavior.
4. Self-dependence--basics of self-help, health, and safety, also including eventual employment and satisfying life in open community.

Much of education in these imperatives is provided during early childhood in the home by parents and others of the family and community. As children enter the schools, a coalition of family and school is formed to advance learning in these essential areas. There is no limit to what individuals can learn in language, mathematics, social skills, and self-dependence. At the upper extreme, one can envision the remarkable individual who is fluent in several languages and in all forms of language expression, who is expert in understanding of basic mathematics and perhaps in international finance, who is well-skilled socially and able to lead in world-class dialogue about social planning, and who is carefully attentive to his/her own health and safety. At the other extreme, one can envision the severely disabled child struggling to master the most basic aspects of communication, social life, and self-help.

The point of importance here is to identify those aspects of learning that are so basic and universally important that we ought to know how everyone is progressing and how effective our school programs are in teaching them. Assessment processes here are not addressed to how well pupils play the piccolo, use the cross-cut saw, or write poetry; such abilities are electives as are literally hundreds of other domains of learning, many of them reflected in school curricula.

The determination of what the imperatives are is a matter of much importance. They will vary among cultures and across time. Just now, for example, it appears that the ability to use a computer is emerging as something near a cultural imperative in our society. One way to identify a cultural imperative is to observe the domains of learning considered to be so essential that the society has established (often at high cost) secondary or "back-up" programs in the schools to give every student a second chance, when needed, for instruction and learning in the area. We have such"back-up" or "second system" programs to help students in exactly the four areas of imperatives listed above. These are the essential topics in most of special education, Chapter I programs, migrant education, and other categorical programs. The obverse is also the case; that is, assessment in domains of the cultural imperatives should be the means for identifying students for whom the current school program is not working well and for whom "second chance" programs are needed.

In summary, it is proposed that assessments ought to be made universally only in the areas of cultural imperatives. This is not to depreciate assessment in other domains, but only to say that assessments should be made in areas of cultural electives only for students who are engaged in studies in these special areas. This is but an application of the principle that assessments should be in accord with the curriculum ("test where you teach"), recognizing that all students should be expected to learn and to be measured for progress in areas of the cultural imperatives.

To assess all students in the imperatives is and will be a considerable technical challenge. For example, it is not feasible to use the same instrument in measuring the reading abilities of all elementary and secondary school students. A number of instruments, varying in level and complexity, would be necessary for such a task. But if one is reasonably clear about what reading ability is and how it generally advances from earliest to highest levels, it is possible to use a series of tools and procedures and then to specify for an entire population of students what levels and rates of achievement are being accomplished.

An example of such an approach might be to examine pupil records for an entire school on a given topic such as general reading ability. Test results of acceptable quality and uniformity may be found for 95% of the pupils. In remaining cases, often involving students enrolled in special education or other categorical programs, it then becomes necessary to use alternative tests, adaptive measurement procedures, and teacher judgments to estimate the reading achievements of each of the remaining individuals and to enter their data with the larger set of data, finally to represent all pupils and the school as a whole. Usually it will be useful to make separate summaries of data by age or grade levels and then to combine all data by appropriate means. In any case, meaningful universal assessments can be made to show the achievements of the entire student body, including literally all students -- even those at the margins.

Often there will be dissatisfaction with the tests or other procedures used in universal assessment practices at a given time. That is more likely in some domains than in others. At this time, for example, it is more difficult to assess progress on social skills and self-dependence than on language and mathematical abilities. There is much work to be done to clarify goals and to improve the technical aspects of assessment in these fundamental areas. The suggested strategy is to start with the realities of a given school in whatever methods of assessment are being used and then to seek the improvement of procedures. Improvements in assessment are sure to be possible now and always.

Who Makes the Decisions About Assessment?

Decisions about assessment can be made at many different levels--national, state, local district, individual school, classroom, or individual student. As a rule, I suggest that decisions about assessment should be made at a level that also has the resources and commitments necessary to follow through in matters of instruction as implicated by the assessments. Also, whoever makes the decisions should be held accountable for the technical adequacy of the procedures to be used. If assessments are made or mandated at a national or state level, but without a parallel provision for support of follow-through programs, the effects tend to be excessively procedural and bureaucratic, causing much aggravated attention but little improvement in learning.

Costs are also a consideration. Universal assessment at a national or state level involves very high expense and the provision at the national level of resources for follow-through programs becomes even more costly. Again, the implication is to be very conservative about broad state and national assessments and to favor local assessments that are closely linked to instructional decisions.

Decisions about assessment and related decision rules can become intensely political, resulting in undue influences upon curriculum and unfair effects upon individual pupils. Broad national and state programs tend especially to be subject to political influences and thus should be developed only very conservatively.

It is proposed that national and state-level universal assessment programs be limited to the domains of the cultural imperatives. That would include assessments in language (probably mainly reading), arithmetic, basic social skills, and self-dependence. Tests might be used in the case of reading and arithmetic. It is an advantage that these are two broadly useful basic skills. They do not imply a favored content. That is, a person who can read may choose freely what he or she reads; and a person who knows how to measure and count may apply those skills to any situation. This puts abilities to read and to count largely beyond political influences.

Assessment in social skills may need to be conducted at least for a time mainly by non-intrusive means such as keeping data on rates of school suspensions and expulsions, delinquency, drug-usage, voting, and the like. In the case of self-dependence, assessments may be made through use of employment statistics, requests for rehabilitation services, and such means. There are important decisions to be made and resources to be provided at state and federal levels concerning programs that affect learning rates in these several domains, for example, special education, back-up programs for students who have difficulty in basic academic skills (see Chapter I program for disadvantaged children), delinquency control, the migrant education program, the various programs for students with "low English proficiency." In effect, authorities earn their right to conduct assessments in these several domains by providing resources and some of the leadership for follow-through instructional programs.

While favoring a very conservative approach covering broad and universalistic assessment programs, there is good reason, I believe, to be more flexible about assessments in areas of cultural electives. The development of high quality assessment procedures (including tests) in areas such as science, geography, history, music, and foreign languages is very important. But assessments in all such areas should actually be made only in the cases of students who are receiving systematic instruction in corresponding areas. And here emphasis is properly given to the upper levels of learning. This is where ideas and practices concerning "world class" performance are relevant.

As noted in later sections of this paper, it is possible to specify "world class" standards in all areas, both imperative and elective, and to specify the rates at which students in the United States (or in particular states) are meeting such standards. But, again, the principle to be observed is that students should face assessments only in domains in which they are also receiving instruction. In the case of "electives" that will not include everyone. There is a current effort to develop national instruments for assessment in a number of domains but the policy being advocated is to leave decisions about using them to the individual states and school systems. That seems wise.

"World Class" Standards: The Consensus Doctorum Principle

In many standard-setting practices, it is common to follow the principle of consensus doctorum. That is, one convenes and seeks consensus among a group of the chief custodians of knowledge in a given field about standards to be applied in that field. That is the way courts decide upon a standard to be held in some matters of dispute. For example, a court may convene experts in metallurgy and ask them about what the standards should be for metal in an automobile axle (perhaps this involves a case of broken axle, an accident, and a personal injury). Also, for example, this is the way accreditation agencies approach problems of standards, as for medical school curricula. Experts are assembled and asked to specify the well-confirmed and important pharmaceutical knowledge, for example, that could and should be taught as part of the medical school curriculum.

Similarly, experts in various curriculum areas (such as science, geography, history, etc.) and in child development can be convened and asked to specify "world class" or other levels of standards for school learning. It would help to have the standards set at a series of development levels, so that there were clearly specified standards in sequence at various age or grade levels. That appears to be exactly the way developments are occurring at a national level at this time. Indications point to standards for students at 4th, 8th, and 12th grade levels.

It is to be recognized that the consensus doctorum procedure is not a totally democratic process. The standards are proposed by an elite group of knowledgeable people and not by a representative group of teachers, parents, or citizens in general. People in general may decide whether they wish to develop and apply standards in a given field, but specifying the standards is the work of experts.

One can imagine that, in the future, one might see reports on schools and students of these kinds:

- Sixty percent of the graduates of Central High School met "world class" standard in at least one curriculum area.
- Twenty-seven percent of the graduates met "world class" standards in geography.
- Fifty-three percent of East Side Elementary School pupils at grade 4 passed the history exam at the "world class" level.
- The median number of "world class" examinations passed by West Side Junior High School eighth graders this year was two.
- The following students of the graduating class at Central High School passed exams at "world class" level in five or more curriculum areas: John Doe, Emily Jones, etc.

It is to be noted that in this treatment of the topic of "world class" standards, the orientation is mainly to domain-oriented assessments. That is, the assessment procedure is very clear as to curriculum areas covered and the processes of comprehension, problem-solving, and creativity expected in each domain. The decision rule is also dichotomous, "pass or not to pass." Not all students need to be assessed in these areas; only those receiving instruction in the area are assessed.

In the cultural elective areas it is not only the schools that make dichotomous decisions. This is the approach made by most institutions. It is decided, for example, that a given person does or does not qualify as a plumber or as a teacher or physician. You get admitted or rejected for enrollment as a student at Harvard. Thus, decision processes and related assessment practices are different for electives than for imperatives. In the case of imperatives, extraordinary efforts are made to include every child. In fact, total inclusion is the law.

Assessment procedures that emphasize high levels of achievement in elective areas and that are carefully attuned to knowledge and skills of emerging importance in our very complex culture are obviously important. Assessments of this level and kind are certainly not to be applied in the case of disabled students who are struggling at the rudimentary levels of the cultural imperatives. There are, of course, some students with disabilities who have remarkable positive abilities and they should by no means be excluded from "world class" assessment procedures. Again, the principle: testing (assessment) and teaching should run in parallel.

Interpreting and Reporting Results of Assessment

The most important level of concerns about results of assessments are at the level of the individual student. The question here is: How does the assessment help the teacher and/or parent understand the individual pupil's skills and views of the world? These are the starting points in arranging an instructional and environmental situation for the child. Assessment at this "clinical" level of the individual is widely treated elsewhere (Salvia & Ysseldyke, 1991), so will not be discussed extensively here; except to note that it can be done in ways that are harmful as well as helpful. There are tendencies, in cases of poor school progress by particular pupils, to move all too readily to assessment processes that ascribe the failure to the individual or "blame the victim," label the child negatively, and isolate or separate the child from the mainstream of the school.

It is common in reporting results of assessments on a class or other aggregate of students to use central tendency statistics only. For example, the average score of students of an entire state or school district may be reported on a reading test or a scholastic aptitude test. The argument here is for a simple but important extension of such reporting procedures to give attention to the margins as well as to the average.

A recently advanced ideal for this is "20/20 analysis" (Reynolds, Zetlin, & Wang, 1992). In 20/20 analysis, one begins by choosing one of the cultural imperatives as expressed in terms of educational outcome. For instance, one may choose to consider data relating to general reading ability of students in a given school. All students are then measured on general reading ability and the 20th, 50th, and 80th percentiles are computed. This makes it possible to identify all pupils whose rates of progress puts them in the lowest one-fifth of their group. Similarly, students in the top 20 percent group are identified.

Reynolds et al. have proposed that 20/20 procedures be used as a first approach in selecting students who most need adaptations in their instructional programs. All students in low-20% status would be studied intensively and plans for alteration in their school programs and life situations would be made collaboratively by parents and teachers.

20/20 analysis is outcomes-oriented and non-categorical in its approach to identifying students who need special help. Prevailing school programs tend to operate exclusively by input-oriented identification procedures and in a narrowly categorical programmatic framework. Eligibility for special education involves a complex and expensive two-step process; first a student is determined to have a disability and then that the student has a "special" educational need. For most exceptional students the first step of the process has no instructional validity; that is, the classifications do not divide treatment (instructional) groups with validity.

It happens, however, that the very simple and low-cost 20/20 procedure identifies most children now served in special education programs. In three elementary schools in a rural Minnesota community it was shown that 100% of special education students identified and placed in special education in traditional ways were identified in the low group in 20/20 analysis (Peterson, Heistad, Peterson, & Reynolds, 1985). In a study conducted in four elementary schools in Utah, it was found that 91% of special education students were identified by using a 20th percentile cut-off on reading test scores (Stone, Curdick, & Swanson, 1988). In two inner city schools of Los Angeles, 79% of special education students were identified by low-20% status in reading. In many schools, virtually all students in low-20 status are enrolled either in special education or Chapter I programs. Looking at the data from another perspective, however, it is also clear that in some schools large numbers of very poor achievers are receiving no special help at all in categorical programs (Reynolds et al., 1992). The fact that many students with major learning problems "fall through the cracks" of categorical programs is considered to be a serious problem.

While proposing that 20/20 analysis might be used as a means of identifying pupils in need of special help, it is acknowledged that other students will need special education. Pupils who are deaf or blind or who have major speech problems, for example, will often need special education even if they achieve in "the imperatives" above a 20th percentile level in their school. Students in high-20 groups often need adapted instruction but here schools most often lack special funding to support "special" programs. That is a matter in need of repair.

The major point here is to propose that results of assessment processes be reported and treated in terms of marginal pupils as well as of central tendencies. In this writer's view, we have too often organized "special" programs in the schools on the basis of input-oriented data and classification/labeling systems that are overly expensive and degrading to students. We can do better by drawing together procedures across categories and with emphasis on both low and high achieving students. 20/20 analysis encourages such a broad, systematic, outcomes-oriented approach to school improvement and in none of the procedures are students labeled and categorized in traditional ways.

Assessment of Programs: Prospects for a Dual Approach

Assessment practices intended to make a practical difference in the lives and learning of children need to attend to programs as well as to students. There is now a voluminous research literature on what characteristics of schools and features of instruction tend to have desirable effects on student learning. Meta-analyses have turned even to meta-reviews, which is to say that massive reviews are now undertaken not just of original studies but of the review literature (Wang, Haertel, & Walberg, 1990). Two kinds of programmatic assessment, both of them illustrating dual approaches, are presented here. By dual approaches it is suggested that often the same data can be used, with only minor adjustments or aggregations, for both individual and program assessment. The first case involves 20/20 analysis and the second shows ways of using variables shown to have positive effects on learning at both individual and programmatic levels.

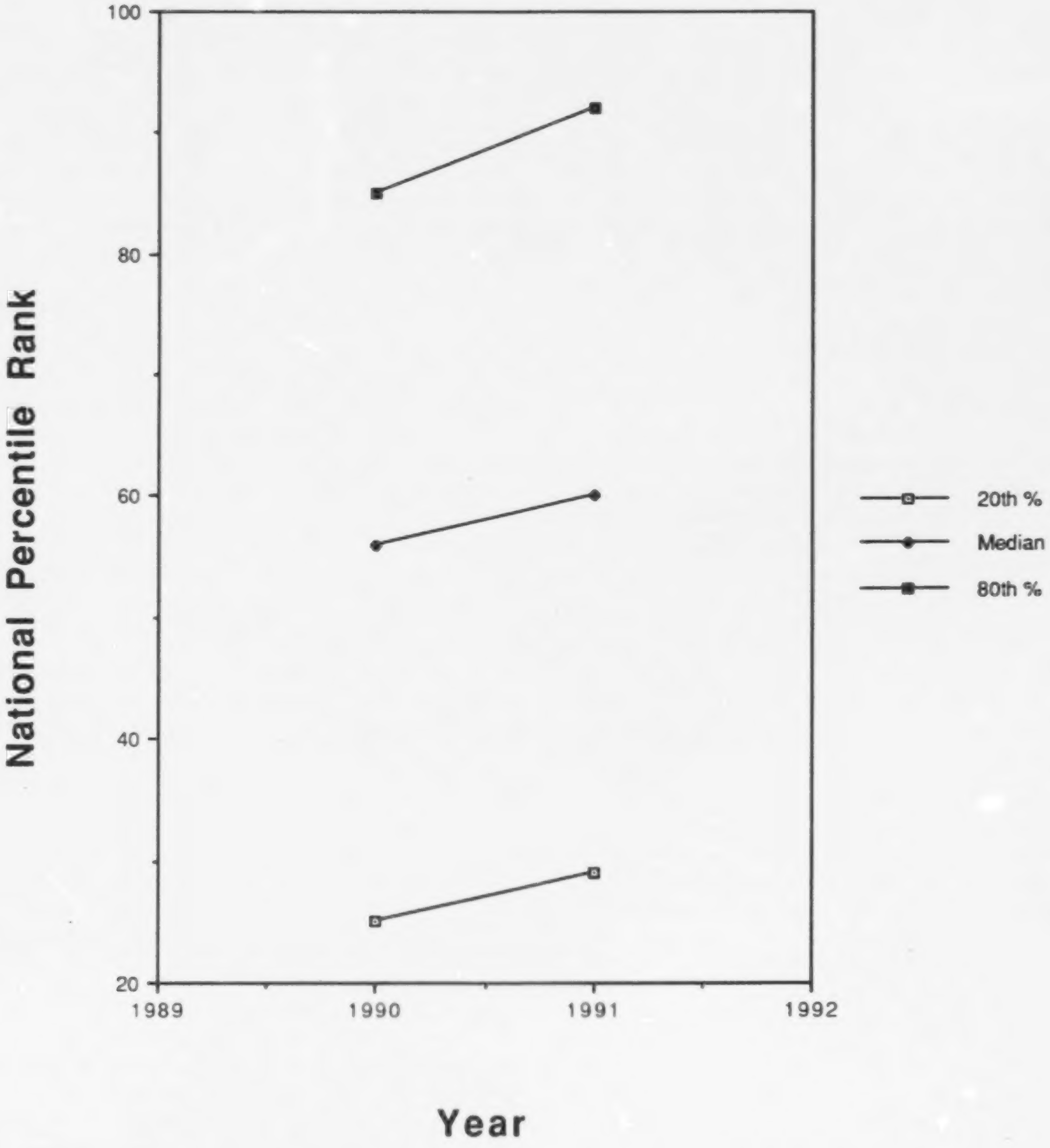
In one large-city school 20/20 analysis was performed, with results for Spring, 1990, showing that the 20th percentile in mathematics at the school was at the 25th percentile nationally. The median at the school in math was at the 56th percentile nationally, and the 80th percentile for the school was at the 85th national percentile. These findings show that the school results were somewhat above national norms both at the margins and at the median (see Figure 1).

A special effort for improvement was made in the school in the mathematics area during the 1990-1991 school year and the results for Spring, 1991, showed favorable results. The 20th percentile for the school moved up to the 29th percentile (national), the median up to the 60th percentile nationally, and the 80 percentile up to the 92nd percentile. Again, see Figure 1. The work in this school over the 1990-1991 school year cannot be described as a true experiment, showing specifically that the interventions used had the specific effects of improved scores on mathematics tests. Nevertheless, by repeated applications of the 20/20 assessment procedures, one can demonstrate with some credibility effects on programmatic outcomes. A major point of emphasis here is on procedures that examine program outcomes at the margins (20th and 80th percentiles) as well as at the "average." Also, the 20/20 data show dual usefulness in identifying individuals at the margins as well as showing aggregate or programmatic outcomes at the margins.

Figure 1

Twenty-Twenty Analysis

CAT Math Concepts/Application Trends



A second approach, again having the feature of duality, is based on the so-called effectiveness literature. The idea is that starting with research evidence on what has effects on learning rates, we can look both at individuals and at programmatic levels. When one can make such a dual approach the introduction of efforts for change in a school or classroom are eased. The total effort is organized around one theme. Consider academic learning time or time on task as an example (see Table 1).

It is well-confirmed that children learn more when they spend more time in learning activities in a given domain. Indeed, nothing is quite so predictive of what and how well a child will learn as the amount of time spent at it. All is for the better, of course, if the time is spent actively and successfully in well-conceived learning activities. Thus, an important consideration of every teacher should be to see that much time, as well as efficient practice, is given in the classroom to the learning of important skills and knowledge. Equally, early consideration should be given to how each individual distributes and uses time when his or her learning is not proceeding well. It may be that extraordinary measures will be required to provide more time for the individual to study in a given field or on a particular set of skills, if better progress is to be made. Some ideas for both the class as a whole and for dealing with individuals are provided in Table 1. Several "principles" taken from the effectiveness literature are listed in the table; the examples of procedures useful with individual pupils and a class-as-a-whole are provided. The point of emphasis here is that assessments of individuals and of the group can be made on the same variables.

Briefly, consider one more example, the second variable listed in Table 1 -- comprehension monitoring. Today learning is viewed not simply as a process of accumulation, gradually adding skills or knowledge. It is seen as transformative, involving reshaped comprehension. The starting point for the teacher in offering instruction is knowing where the student is now -- what views the student has of the world. Teaching is designed to help students make sense of what they know. This suggests that we need procedures for probing the pupils' understandings or comprehension. When a pupil has fallen behind, it is hard for the busy regular class teacher to know what is in each student's head, what the student comprehends, and it is easy to neglect the student. Teachers can go into a problem-minimizing approach (Scardamalia & Bereiter, 1985), such as:

- sending the student away, as in referral to a special education program
- just eliminating content or simplifying the curriculum excessively
- focusing just on interest and enjoyment
- engaging in decontextualized skill teaching; "just do the work sheet"
- "covering the subject matter" without comprehension checks
- neglecting practical applications/becoming text-bookish
- training on tasks that fall short of goals (e.g., do remote perceptual training, assuming transfer to reading skills)

Avoiding such retrenchments in the education of the low-achieving pupil and moving instead in an aggressive individualized way is what is needed, but that will often require the extra help of a teacher who has time and talent for understanding and teaching the individual. Thus, there is need for detailed careful assessment of the individual student's comprehension, both as to background and as instruction proceeds. By aggregating such data, one has a basis for group activities as well. I assume that the remainder of Table 1 is self-explanatory. Additional variables could be considered, but not in this paper.

Table 1

Illustrative Procedures for Effecting Learning Rates

Illustrative Procedures for the Class as a Whole	**Illustrative Procedures for the Slow-Progress Individual Pupil**
Principle: Increase Active Learning Time; Time on Task	
• Provide efficient management of transitions between class activities • Teach for efficient, time-saving class routines; do so early in school year and be consistent • Use efficient, clear start-up plans for class each day • Clear out all possible interruptions of instruction • Establish clear alternative activities for use by students when assignments are complete • Extend study and practice through well-designed homework (thoroughly checked) • Assess periodically the use of time in the class and make appropriate changes in time allocations • Stretch time given to basic skills--as in reading and arithmetic	• Schedule extra time for instruction on critical subjects and skills • Reinforce increases of sustained attention and effort • Provide massive high success experiences in important learning areas in drives for automaticity on basics • Allocate added time through Summer programs • Try for extended study and practice at home • Assess the individual frequently on use of time in school, checking also for what produces improvements • Use peer-mediated coaching to extend time and effort in studies

Class as a Whole	Slow-Progress Individual Pupil
Principle: Monitor Student Comprehension Frequently	
• Use domain-referenced tests to check on comprehension of students in specific topic and skill areas • Use questions in class to check on common misperceptions and student comprehension at all stages of instruction • Model (by teacher) ways of self-monitoring of comprehension	• Probe in detail to understand comprehension and misconceptions of the individual in domains of instruction • Teach aggressively for comprehension in one-on-one or small-group situations • Teach individual methods of self-monitoring for comprehension • Provide close and continuous monitoring of student performance as a means of educational diagnosis and planning • Provide quick daily reviews of material covered the day or week before • Seek parent cooperation in monitoring pupil progress on specific objectives and goals
Principle: Provide a Safe, Orderly School Environment	
• Arrange the classroom so as to support good order and efficiency • Teach essential rules, procedures, and expectations for the class and school; do it early • Demonstrate "withitness" (the teacher understands all that occurs in the classroom) • Use disciplinary procedures consistently and with fairness • Clarify expectations for classroom and school behavior • Teach students to participate in management of classroom routines	• Repeat and reinforce teaching on essential class rules and procedures • Clarify and teach for careful recording of class assignments and expectations • Reinforce (reward) individual progress in orderliness • Teach specific social skills that will enhance positive relationships with classmates and teachers • Arrange peer assistance to assure on-time and orderly behavior

Class as a Whole	Slow-Progress Individual Pupil
Principle: Provide Frequent Feedback to Students About Their Performance	
• Return all tests and discuss results • Encourage and assist students in monitoring their own progress in learning • Check all homework • Comment frequently on progress and common sequences in learning in the domain of instruction • Provide models of desired performance and of performance analysis • Arrange materials, sequences, and expectations with due consideration of prior learning, so that most feedback is positive	• Use individual "charting" to give student a record (an "external clock") of learning progress • Teach students to sense and to "chart" their own individual progress in learning • Review prior learning systematically, noting individual progress • Use reinforcements principles systematically to facilitate desired learning and to extinguish undesirable behavior
Principle: Make Learning Tasks Appropriate in Difficulty	
• Provide a range of reading materials and other instructional tools, taking into account individual differences in specific skills and general knowledge of students • Measure specific background of students on topics of the curriculum as an aspect of planning instruction (e.g., check vocabulary and major concepts)	• Measure individual skills and abilities carefully in domains relevant to instruction • Make appropriately adaptive arrangements in reading and other instructional materials, so that the individual has mostly success experiences • Use compensatory approaches (e.g., use tape-recorded materials when student cannot read existing materials) when necessary to avoid delays in curriculum experiences for an individual • Move student to advanced topics only when essential prerequisites are mastered

Class as a Whole	Slow-Progress Individual Pupil
Principle: Promote Self-Responsibility, Meta-Cognitive Learning Strategies and Motivation for Continued Learning	
• Provide models of meta-cognitive approaches to study and problem solving • Give students gradually increased opportunities to make decisions about how they organize their own learning activities • Hold definite standards for prompt, high-quality performance in school-assigned tasks • Discuss long-term and career-oriented aspects of each topic of the curriculum in ongoing ways • Teach students to measure their own performance and to use data in planning their own development • Permit students to select some learning goals and activities	• Teach specifically and individually for self-awareness and planning of study activists • Assist individual student in keeping track of assignments and schedules for school work; work toward self-responsibility • Provide experience for the student in clarifying one's own goals for study activities • Teach for flexibility and appropriateness in the individual's approach to reading and other tasks
Principle: Work for Positive Attitudes Toward Schools, Learning, and Teachers	
• Recognize and reward outstanding performance in desired areas of learning • Use cooperative groups in appropriate aspects of instruction and teach for effective group behavior • Encourage students to seek help from peers and to provide help to classmates • Provide recognition and rewards for students who make special contributions to positive feelings among students and about school and who help to create mutual trust among students and teachers	• Reinforce in individual sessions skills of cooperative group work • Arrange peer contacts that foster positive attitudes toward learning and schools • Teach specific prosocial skills to student showing social skill problems

Class as a Whole	Slow-Progress Individual Pupil
Principle: Use Clear, Organized, Direct Instruction	
• Take time to link important ideas as instruction proceeds • Uses advance organizers, content overviews, and reviews of objectives to give students clear and integrated perceptions of the curriculum • Pace presentations rapidly • Sequence information carefully • Engage students in active, frequent responding to show skills and comprehension	• Use high density, highly structured teaching in one-on-one or small group situation to • Micro-teach with direct, frequent feedback on topics chosen through diagnosis of individual's skills and knowledge • Maximize teacher-directed learning activities on carefully sequenced topics
Principle: Maintain Clear Expectations of Content Mastery	
• Provide for systematic reviews of past learning and make students aware of learning progress • Take time to specify instructional goals for students • Provide models of excellent performance • Show how one can read or engage in other academic work in different ways for different purposes	• Use highly precise ("pinpointed") objectives for individual pupil and direct instruction to a definite sequence of such objectives • Involve student in assessments and decisions about content mastery and movement to more advanced topics
Principle: Encourage Parents to be Supportive and Involved in Students' School Work and to Hold High Expectations for Academic Success	
• Keep parents well informed about curriculum content and objectives • Involve parents in school work as volunteer assistants and advisors	• Prepare highly educational plans for individual student cooperatively with parents • Review individual student progress with parents on regular schedule • Write "contracts" with parents covering mutual commitments to teach and support individual student

Fortunately, a variety of tools are now becoming available for use in assessing variables such as those listed in Table 1 at both individual and programmatic levels. For example, see Ysseldyke and Christensen (1987), and Deno and Mirkin (1977). In this writer's view, the educational diagnosis of individual pupils and of school programs could be enormously improved by concentrating attention on principles and variables, such as those listed in Table 1. These are among the variables that have been demonstrated to have positive effects on the learning outcomes; and they are directly in control, in at least some degree, by teachers. By the duality feature they bring attention to both individual students and school programs, which should help to dispel the common arguments about who or what is to be blamed for educational failure. Instead, it focuses on what must have attention--both in pupil behavior and programmatic features -- to cause improvements in learning. It could be transformative in special education if attention were to shift to variables that truly make a difference in the life and learning of children.

A Brief Look to the Future

The problems and issues of assessment in the schools will not be solved quickly or soon. Hopefully, there will be continuous improvements, based on new insights and careful tests of new ideas. One idea for the future is proposed here.

It is that we learn to measure time and make time-for-learning the main variable used to reflect individual differences. This is not a new idea (Carroll, 1963); but a practical approach, applicable to exceptional students, has been proposed recently (Reynolds, Heistad, Peterson, & Dehli, 1992). It involves numbering school days continuously through the grades. In grade 1, day numbers might range from 1 to about 180; then in grade 2, from about 181 to 360; in grade 3 from 361 to 540, etc. All exams and other written school work would carry the day number on which each piece was completed.

The "days to learn" system was used in a small set of elementary schools that followed a common curriculum and used the same set of mastery exams. Among many other findings, it was shown that completion of the 10th unit in mathematics required 92 school days for pupils not enrolled in any categorical program. To reach the same level in mathematics, special education students required an average of 131 days or about 7 weeks more of instruction than other pupils.

As domain-referenced assessment procedures gain in use it appears likely that the variable of time (such as in "days to learn") will be useful in showing developmental trends. How many days did it take for the median student in the school to pass the domain-referenced exam on combinations of two-place numbers? How early in the primary grades did pupil x show extraordinary difficulty (in terms of time to learn) in the reading curriculum? In 20/20 analysis, using time for learning as the variable, who were the students below the 20th percentile and above the 80th percentile?

The time variable has characteristics that are quite uncommon in education. It has ratio scale properties. That is, 10 days are twice as many as 5 days in a very meaningful sense. It costs twice as much to employ a teacher for 20 days as for 10 days. We can show that students receiving individual tutoring make progress (or do not make progress) at some precise rate as compared with rates of progress without such tutoring. Time is the most ubiquitous variable in educational research on what effects learning. It seems inevitable that we will learn to measure time and to use it increasingly in future assessment activities.

Summary

It has been proposed that universalistic assessment (including every student) be limited to domains of the cultural imperatives, defined as the skills and knowledge essential to everyone in a given society. For the United States at this time it is suggested that the cultural imperatives fall into four general areas: language, mathematics, social skills, and self-dependence. In other domains, the cultural electives, it is proposed that assessments be conducted only in areas in which students are receiving instruction. Since not all students study in the same elective areas, assessments would, in this case, not be universalistic. Only some students would be assessed in such cases. An approach to issues concerning so-called "world class" standards is proposed, following the principles of consensus doctorum. New ways of reporting results of assessment that are oriented to "world class" standards are presented. It is proposed that individuals and school programs be assessed in parallel fashion and by dual procedures. In deciding upon variables to be entered into dual assessments it is proposed that priority be given to major outcomes sought in the instructional programs of the school and to variables known to have positive effects on learning outcomes. In reporting results of assessment activities it is urged that data be presented that reflect performance of students at the margins as well as at central positions. The case of "20/20 analysis" is presented as an example of how one can look to the margins. In all of this a case is defined for including "disabled" students, indeed literally all students, in assessments in areas of the cultural imperatives. And it is argued that most exceptional students can and should be identified by using assessments in the domains of the cultural imperatives, rather than by traditional means of classifying, categorizing, and labeling. In a brief look to the future, it is anticipated that time-for-learning will become a chief variable in measuring learning rates and in working for improved learning.

References

Carroll, J. B. (1963). A model of school learning. Teachers College Record, 64, 723-733.

Deno, S. L., & Mirkin, P. K. (1977). Data based program modification: A manual. Reston, VA: Council for Exceptional Children.

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). Placing children in special education: A strategy for equity. Washington, DC: National Academy Press.

Hively, W., & Reynolds, M. C. (Eds.) (1975). Domain-referenced testing in special education. Reston, VA: Council for Exceptional Children.

Koretz, D. M., Madus, G. G., Haertel, E., & Benton, A. E. (1992). National educational standards and testing: A response to the recommendations of the National Council on Education Standards and Testing. Santa Monica, CA: RAND.

McGrew, K.S., Thurlow, M.L., Shriner, J.G., & Spiegel, A.N. (1992). Inclusion of students with disabilities in national and state data collection programs (Technical Report 2). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

National Council on Education Standards and Testing. (1992). Raising standards for American education: A report to Congress, the Secretary of Education, the National Goals Panel, and the American people. Washington, DC: Author.

Peterson, J., Heistad, D., Peterson, D., & Reynolds, M. C. (1985). Montevideo individualized prescriptive instructional management system. Exceptional Children, 52 (3), 239-243.

Reynolds, M. C., Heistad, D., Peterson, J., & Dehli, R. (1992). A study of days to learn. Remedial and Special Education, 13 (4), 20-26.

Reynolds, M. C., Wang, M. C., & Walberg, H. J. (1992). The knowledge bases for special and general education. Remedial and Special Education, 13 (5), 6-10.

Reynolds, M.C., Zetlin, A., & Wang, M. C. (1992). 20/20 analysis: Taking a close look at the margins. Exceptional Children, 59(4), 294-300.

Salvia, J., & Ysseldyke, J. E. (1991). Assessment. Boston: Houghton Mifflin.

Scardamalia, M., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S. Chipman, W. Segal, & R. Glaser (Eds.), Thinking and learning skills: Research and open questions (Vol. 2) (pp. 563-577). Hillsdale, NJ: Erlbaum.

Sinclair, R. L., & Ghory, W. J. (Eds.). (1987). Reaching marginal students: A primary concern for school renewal. Chicago: McCutchan Publishing Corporation for the National Security for the Study of Education.

Stoddard, G. (1961). The dual progress plan. New York: Harper & Bros.

Stone, B., Curdick, B. P., & Swanson, D. (1988). Special education screening system: Group achievement test. Exceptional Children, 55 (1), 71-75.

Tyler, R. W. (1987). Marginality in schools. In R. L. Sinclair & W. J. Ghory (Eds.), Reaching marginal students: A primary concern for school renewal. Chicago: McCutchan.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A context analysis of review literature. Journal of Educational Research, 84 (1), 30-43.

Ysseldyke, J. E., & Christensen, S. L. (1987). The instructional environment scale (TIES). Austin, TX: Pro-Ed.